

Qy	1	MKPLAI PANHVCMQOEKHSI.PADFKI.LHTDLSLHPQYTHVSSSHSGCSTTSBGSSSL	60
Db	1	MKPLAI PANHVCMQOEKHSI.PADFKI.LHTDLSLHPQYTHVSSSHSGCSTTSBGSSSL	60
Qy	61	DIYQATESEAGDMDSLGLPETAVDSEDDDDEEDIERASDPLMSRDYIRDCLEKPIDRTD	120
Db	61	DIYQATESEAGDMDSLGLPETAVDSEDDDDEEDIERASDPLMSRDYIRDCLEKPIDRTD	120
Qy	121	DDIQLLEFHMQLPAFANMTMSVSRRELCAVWFAVVERAGITV.LNDGEELDSWVILNGS	180
Db	121	DDIQLLEFHMQLPAFANMTMSVSRRELCAVWFAVVERAGITV.LNDGEELDSWVILNGS	180
Qy	181	VEVITYPDGKAEILCMGNSFGVSPITMDXEYMGKVMRTAVDCQFVCIADQDYCRILANVER	240
Db	181	VEVITYPDGKAEILCMGNSFGVSPITMDXEYMGKVMRTAVDCQFVCIADQDYCRILANVER	240

```
QY 241 NMOKVEEYVWKEHRELDRTGRKHIIYIKGTSERLTMLVEHSHVYDPTFIEDLL 300
DB 241 NMOKVEEELIYVWKEHRELDRTGRKHIIYIKGTSERLTMLVEHSHVYDPTFIEDLL 300
QY 301 TYRTFLSSPMEVGGKLLWENPDLSDKVTYVLLVMVNNHENDFEGDPAWTFLEEFENN 360
DB 301 TYRTFLSSPMEVGGKLLWENPDLSDKVTYVLLVMVNNHENDFEGDPAWTFLEEFENN 360
QY 361 LEREMGGLRLNLTAACAARRLMTLTKPSREPPLPILLGSEKGFIVDSVDSGS 420
DB 361 LEREMGGLRLNLTAACAARRLMTLTKPSREPPLPILLGSEKGFIVDSVDSGS 420
QY 421 KATEAGLKGQDQILEVNGCFENFIOLSKAMELIRNNTHTSIVKTNLVEFKELRLSEE 480
DB 421 KATEAGLKGQDQILEVNGCFENFIOLSKAMELIRNNTHTSIVKTNLVEFKELRLSEE 480
QY 481 KNGAPHLPKJIDIKKASRYSLPDLAVDEQVIGLEKYNKSKANTVGGRNKKILDKT 540
DB 481 KNGAPHLPKJIDIKKASRYSLPDLAVDEQVIGLEKYNKSKANTVGGRNKKILDKT 540
QY 541 RSLIPQKPYNDIGIGSODDSIVGLRQTKHPTALPVSGTLSSNPDLQSHHRLDFS 600
DB 541 RSLIPQKPYNDIGIGSODDSIVGLRQTKHPTALPVSGTLSSNPDLQSHHRLDFS 600
QY 601 ATPDLPDVLRVFKADQDSRYIMISKDTAKEVYIOALREFAVTATPPQYSLCEVSVTPE 660
DB 601 ATPDLPDVLRVFKADQDSRYIMISKDTAKEVYIOALREFAVTATPPQYSLCEVSVTPE 660
QY 661 GVIKORRLPDOLSKLADRIQLSGRYLYLKNMETETLCSDEDAQELLRESQSLQLSTVE 720
DB 661 GVIKORRLPDOLSKLADRIQLSGRYLYLKNMETETLCSDEDAQELLRESQSLQLSTVE 720
QY 721 VATOLSMNFELEFRNIEPTETIDOLFRLKSTSCANLKRFEVINOEFVWASLRETN 780
DB 721 VATOLSMNFELEFRNIEPTETIDOLFRLKSTSCANLKRFEVINOEFVWASLRETN 780
QY 781 QLRKWKIKHFKIALHCRECKNFNSMFAISGLNLAVARLRTTWELPKRYEKLFDL 840
DB 781 QLRKWKIKHFKIALHCRECKNFNSMFAISGLNLAVARLRTTWELPKRYEKLFDL 840
QY 841 ODLDPSPNNAKRYNVLNSQNLQPPITPLFPVYIKDLTFLEHGNDSKVDGLVNEFKLMI 900
DB 841 ODLDPSPNNAKRYNVLNSQNLQPPITPLFPVYIKDLTFLEHGNDSKVDGLVNEFKLMI 900
QY 901 AKETIRHGRMASVMDPALMFRTRKKKWRSLGSLSGSTNATVLDVAQTGCHKRRVRSS 960
DB 901 AKETIRHGRMASVMDPALMFRTRKKKWRSLGSLSGSTNATVLDVAQTGCHKRRVRSS 960
QY 961 FLNAKKLYEDAQARKVQYLSNLELMEDEESLQTLSDCEPATNTLPKNBGDKKPVKSE 1020
DB 961 FLNAKKLYEDAQARKVQYLSNLELMEDEESLQTLSDCEPATNTLPKNBGDKKPVKSE 1020
QY 1021 TSPVAPRAGSQQAQSLPQOQOQPPRAHKINOGIQVAVSLYPSRKKVPVKDLPEGINS 1080
DB 1021 TSPVAPRAGSQQAQSLPQOQOQPPRAHKINOGIQVAVSLYPSRKKVPVKDLPEGINS 1080
QY 1081 POLAKKILSLSEEGSLERHKKQAEEDTISNASSQLSPSPQSPPRGYTLABGVTDNF 1140
DB 1081 POLAKKILSLSEEGSLERHKKQAEEDTISNASSQLSPSPQSPPRGYTLABGVTDNF 1140
QY 1141 SDSGHSSEISSRSIVNSFSFDSVPSLHDERORHVSIVETNLGMRMERTMIEPDY 1200
DB 1141 SDSGHSSEISSRSIVNSFSFDSVPSLHDERORHVSIVETNLGMRMERTMIEPDY 1200
QY 1201 SLGSAVMSGRCGLXATATYISSPTELSQDQGRASLDAASGSGSWTSCSGSHDNT 1260
DB 1201 SLGSAVMSGRCGLXATATYISSPTELSQDQGRASLDAASGSGSWTSCSGSHDNT 1260
QY 1261 OTIOHQRSWETLFPFGHTHFYSGDPALMASSSHMDIMFSDSTXKXNRNOSRESLEA 1320
DB 1261 OTIOHQRSWETLFPFGHTHFYSGDPALMASSSHMDIMFSDSTXKXNRNOSRESLEA 1320
```

```
QY 1321 QSRASWASTGYWGEDSGDTGTTIKRGGKDVSLTEASSSLTSTVTEETKRVPMPAHIAV 1380
DB 1321 QSRASWASTGYWGEDSGDTGTTIKRGGKDVSLTEASSSLTSTVTEETKRVPMPAHIAV 1380
QY 1381 ASSTTKGLIARKEGRYRPPPTPGYIGIPIITDPECHSHPARPPYUNALORSRVAR 1440
DB 1381 ASSTTKGLIARKEGRYRPPPTPGYIGIPIITDPECHSHPARPPYUNALORSRVAR 1440
QY 1441 SSDTAGPSSVQOPGHPTSSRPVANKPQWHPKNESDPRLAPYOSQGFSTEDEDEQVSAY 1499
DB 1441 SSDTAGPSSVQOPGHPTSSRPVANKPQWHPKNESDPRLAPYOSQGFSTEDEDEQVSAY 1499
```

```
RESULT 2
US-09-911-826A-7
; Sequence 7, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; FILE REFERENCE: DDW-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
; PRIOR FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-911-826A-7
```

```
Query Match 10.1%; Score 777.5; DB 10; Length 307;
Best Local Similarity 51.3%; Pred. No. 8.2e-40;
Matches 157; Conservative 59; Mismatches 77; Indels 13; Gaps 5;
```

```
QY 709 SQSLSLQSLSTEVEATQLSMNFELFRNIEPTETIDOLFRLKSTSCANLKRFEVINOET 768
DB 1 SNVHFLHNLNAYELAQTLQDQDFANFRQIESTEVYDELFLERKSYGVPMLSKFAELVNRM 60
QY 769 FVWASELFRNTDLKRMKIKHFKIALHCRECKNFNSMFAISGLNLAVARLRTTWEL 828
DB 61 FVWASELCAEHNTVRKIKVQFIKIALHCRECKNFNSMFAISGLHGAVERLQRTWEL 120
QY 829 LPNKYERLFQDLDLPDPSRNNAKRYNVLNSQNL-QPPIIPLFPVYIKDLTFLEHGNDSK 887
DB 121 LPKRYERLFQDLDLPDPSRNMSKYQOLVSAELLAQHPPIIPLFPVYIKDLTFILHGNDR 180
QY 888 VDLGVNFEKLRMLAKIRHVGMAVNMND-PAIMFRTRKKKWRSLGSLSGSTNA-- 941
DB 181 VDLGVNFEKLRMLAKIRHVGMAVNMND-PAIMFRTRKKKWRSLGSLSGSTNA-- 941
QY 942 -TYLDAQTGGH--KKVRRSSEFLNAKKLYEDAQARKVQYLSNLELMEDEESLQTL 997
DB 241 GTV--IAANAGATIRKRRKSTAAIPMKMFEBQAVRRKAVLNSKLISDEDILHKFS 298
QY 998 LOCEPA 1003
DB 299 LECEPA 304
```

RESULT 3
US-09-911-826A-4
; Sequence 4, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; TITLE OF INVENTION: Methods of Use

	Query Match	6.28;	Score 475.5;	DB 10;	Length 261;
	Best Local Similarity	38.98%	Pred. No. 1.1e-21;		
	Matches 119; Conservative	42;	Mismatches 82;	Indels 63;	Gaps 8
OY	713 LLOLSTVEAYDLSMRNFELFRNIEFTETIIDDLFKIRSTSCA-----NIKREEVINOE	767			
	: : : : : :				

Db 1 LLLDPEELAEQLLLDELFRKIEPSSELLGSGVMSKRSKSPPLAQNLEAFIERENEV 60
QY 768 TFWVASILLRET--NQLRMKIKHFIKIALHGRECKNFNSMFAITSLNAPARLRTT 825
Db 61 SNWVATELTKQTTLKPKRAEVLSEFIVAKHCRELNFNNSLMAIVSALSSPSIRLKT 120
QY 826 WEKLPNKYEKLEFQDLQDLFDPSS-RNMAYRNVL-----NSQNLQPIIPLFVYIKDLTFL 880
Db 121 WEKLSKRYKLFEELEELLDPSERNFKYREALKSCNKSPPVQPCVDFGLGYIKDLTFL 180
QY 881 HEGN-----DSKVDGLVNEKILRMATAKLRHVGRAVSVMDDPALMFRTRKKKWSLSLSQ 936
Db 181 DEGNDFLENGTGLVNEKRRKIAKILREI-----RQLQSACQ 219
QY 937 GSTNATVDAVQGTGHHKRVRRSSFLNAKKLYEDQAMARKYKOYLSNLEMEDESIQTL 996
Db 220 P-----YMLKPRNDIQELLRASPLLEVPEBEDE--LYEL 253
QY 997 SLOCEP 1002
Db 254 SLRIEP 259

RESULT 6
US-09-940-836A-6
Sequence 6, Application US/09940836A
Patent No. US20020146800A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Curtis, Rory
TITLE OF INVENTION: 48921, A NOVEL HUMAN GTP RELEASING
TITLE OF INVENTION: FACTOR AND USES THEREFOR
FILE REFERENCE: 38155-20031.00
CURRENT APPLICATION NUMBER: US/09/940,836A
CURRENT FILING DATE: 2001-08-27
PRIOR APPLICATION NUMBER: US 60/228,760
PRIOR FILING DATE: 2000-08-30
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 246
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus amino acid sequence
US-09-940-836A-6

Query Match 6.1%; Score 474.5; DB 10; Length 246;
Best Local Similarity 45.8%; Pred. No. 1.2e-21;
Matches 97; Conservative 39; Mismatches 65; Indels 11; Gaps 3;
QY 707 RESQISLLQSLVYEVATQLSMRNFELFRNIEPEYIDDLFKLSKTS---CANLKREEE 762
Db 4 RKVSLDILDLDEELAEQLLLDELFRKIEPSSELLGSGVMSKRSKSPPLAQNLEAFIERENEV 60
QY 763 VINOETFWVASILLRETQNLRMKIKHFIKIALHGRECKNFNSMFAITSLNAPARLRTT 825
Db 64 RENNISYVATELSEKQQAQAKIIEKFIKVAOHCRELNFNSLMAIVSALSSPSIRLKT 123
QY 823 RTTWKLPNKYEKLEFQDLQDLFDPSS-RNMAYRNVL-----NSQNLQPIIPLFVYIKDL 877
Db 124 KTWKLVKPEKTFEELSELMDPSNNYKNYRELKSCNKSPPVQPCVDFGLGYIKDLTFL 180
QY 878 TFLHEGNDKVD--GLVNEKILRMATAKLRHV 907
Db 184 TFIHEGNPDYLDNTNLINFEKRMIAKILREI 215

RESULT 7
US-09-922-199A-9
Sequence 9, Application US/09922199A
Publication No. US20020187138A1
GENERAL INFORMATION:

APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 15368, A NOVEL HUMAN GTP-RELEASING
TITLE OF INVENTION: FACTOR FAMILY MEMBER AND USES THEREFOR
FILE REFERENCE: 381552002500
CURRENT APPLICATION NUMBER: US/09/922,199A
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 60/222,622
PRIOR FILING DATE: 2000-08-02
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 238
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus amino acid sequence
US-09-922-199A-9

Query Match 6.1%; Score 468.5; DB 9; Length 238;
Best Local Similarity 46.8%; Pred. No. 2.6e-21;
Matches 96; Conservative 37; Mismatches 62; Indels 11; Gaps 3;
QY 713 LQSLVYEVATQLSMRNFELFRNIEPEYIDDLFKLSKTS---CANLKREEYINQET 768
Db 1 LLLDPEELAEQLLLDELFRKIEPSSELLGSGVMSKRSKSPPLAQNLEAFIERENEV 60
QY 769 FFWASILLRETQNLRMKIKHFIKIALHGRECKNFNSMFAITSLNAPARLRTTWEK 828
Db 61 YWVATELSEKQQAQAKIIEKFIKVAOHCRELNFNSLMAIVSALSSPSIRLKTWEK 120
QY 829 LPNKYEKLEFQDLQDLFDPSS-RNMAYRNVL-----NSQNLQPIIPLFVYIKDLTFLHEG 883
Db 121 VKETKRLFEELSELMDPSNNYKNYRELKSCNKSPPVQPCVDFGLGYIKDLTFLHEG 180
QY 884 NDSKVD--GLVNEKILRMATAKLRHV 907
Db 181 NPDYLDNTNLINFEKRMIAKILREI 206

RESULT 8
US-09-911-826A-8
Sequence 8, Application US/09911826A
Patent No. US20020143164A1
GENERAL INFORMATION:
APPLICANT: Rotin, Daniela and Pham, Nam
TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: DDM-5001-US
CURRENT APPLICATION NUMBER: US/09/911,826A
CURRENT FILING DATE: 2002-02-26
PRIOR APPLICATION NUMBER: PCT/CA00/00042
PRIOR FILING DATE: 2000-01-20
PRIOR APPLICATION NUMBER: 2,259,830
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 270
TYPE: PRT
ORGANISM: Homo sapiens
US-09-911-826A-8

Query Match 5.2%; Score 399; DB 10; Length 270;
Best Local Similarity 32.1%; Pred. No. 4.9e-17;
Matches 98; Conservative 52; Mismatches 95; Indels 60; Gaps 5;
QY 716 LSTVEVATQLSMRNFELFRNIEPEYIDDLF---KLSKTSKANLKRFEVINOETFWVA 772
Db 8 VSARDLAGQLTDHDSLFNSIHQVELHYVLPQHLDVTT-ANLEFRMRFRNELGYWA 66
QY 773 SEIRETNQNLKRMKIKHFIKIALHGRECKNFNSMFAITSLNAPARLRTTWEKLPNK 832
Db 67 TELGLCEVPGPRAQLLKKFIKLAHLEKEQKLNLSFPAVMFGLSNSAISRLAHTWRLPKR 126


```

Db 284 MIEFIDVARECFNIGNENSLMIAIISGNMSPVSRLLTKTAKV--KRAK-  

QY 847 SRNMAKYRNVLS-----QNLQPIIPIPLFVYIKDLTFLHEGDSKV-DGLVFEK  

Db 341 SSNFNYNPTALRGAAQSRSTLASHSSREKIVIPFSLIKDIYFLNCGANRLPNCHINFEK  

QY 897 LRLMIAKELRHHGKRAVSVMMDPALMFRTIRKKKWRSLGSLSGCSTNATVLDVAQTCGHEKKRV  

Db 401 FWEIAKQV-----408
QY 957 RRSFLNKKKLIEDAQAARKKQVQYLSNLELMEDESIQTLSLQCEPATINLPKN 1010  

Db 409 --SEPMWKQVCEPPEPRDRKTLQTLVLPV-FSDADVLAYLSEEGENNIEKD 459

```

RESULT 12
 US-09-765-298A-18
 : Sequence 18, Application US/09765298A
 : Patent No. US20020137017A1
 : GENERAL INFORMATION:
 : APPLICANT: AROHNHEIM, AMI
 : TITLE OF INVENTION: METHOD FOR DETECTION PROTEIN-PROTEIN INTERACTIONS AND A KIT THERE
 : FILE REFERENCE: 108387_01
 : CURRENT APPLICATION NUMBER: US/09/765,298A
 : CURRENT FILING DATE: 2001-01-22
 : PRIOR APPLICATION NUMBER: IL 125456
 : PRIOR FILING DATE: 1998-07-22
 : PRIOR APPLICATION NUMBER: IL 128017
 : PRIOR FILING DATE: 1999-01-12
 : NUMBER OF SEQ ID NOS: 31
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 18
 : LENGTH: 684
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 :S-09-765-298A-18

Query Match	3.7%	Score 283;	DB 10;	Length 684;
Best Local Similarly	22.1%	Pred. No. 1.9e-09;		
Matches 122; Conservative	82;	Mismatches 188;	Indels 160;	Gaps 20;

Qy	713	LLOLSTVEVATNOLSMRNFELPRNTEPTREYIDDLKFLRSK-TSCANLKRFEVYNQETPMV	771
Db	127	LUTLHLEIARQOLTTLESDDLRAQVPELVGSWYKEDKINSPLNKLMIHRTTNLTLMF	186
Qy	772	ASELLRETNOLKRMKIKIHKFIKIALHORECKNFENSMFAIISGLNAPVALRTTWELPN	831
Db	187	EKCIYETENLEERVAAYVSRIEILQVQELNFGVLEEVANSANSPYVRDLHTEFQIS	246
Qy	832	KYELTFODLODLPR--SRMAKRYANVLNSQNIOPRIPLFPVPIKQDLFTLHEGNSKYD	889
Db	247	RQKILDEAHLESLDHYKKYLAKLRSI-----NPPCVPEFGIYLTNLIKTEGNC----	296
Qy	890	GLVFEELRMIAKEIRHVGHMASVNMNDPALMFRTRKKWMSLQSLQSOGSNATVLVAQT	949
Db	297	-----EVLKHKHCKL-----INFSKRRVAITETEIQYQONQPCLVNES-	336
Qy	950	GCHKRRVRSSEFLNMAKLYEDA-----QMARVKYOLSLNLEMEDESLQTLSSQEPAT	1004
Db	337	-----DIKFFENLMPGMSMEKETFDYDLFNKSLTEIEBRN-----P	372
Qy	1005	NTLPRKNBGDKK--PVKSE-TSPVAPRAGQQAQSLPPOQOOPRAHKINQGLQVPAYS	1066
Db	373	KPLPRFP--KKYSIPLKSPGVRSNRPFGIMRH-----PPLPODEP-----	411
Qy	1061	LYPSRKVYPVKDLPPFGINSQALKILSLSEBGSLEHRHKKQAEDTISNASS--OLSSP	1117
Db	412	-----RKISRSRIP-----ESFETSPASAPNSRPRLTPR	441
Qy	1118	PTSPQSSPRKGYTLAPSGTYDNFS--DSGHS-----EISSKSSST	1155
Db	442	PAGASS-----TTDVCVSFDSHDSFSSHSNDTVFIQVTLPHGPRASAVSSTIS	490

```
OY 1156 SNSSPDVSPLVTLHDERORHSIVETN--LCGMGRERTMIEPDQSLGSAVMASGR 1212Z
      ||| : : : : : |
Db 491 LTKGVDEPVPPVPVPRRFEESAPAESPSKIMSHLDPALPPROPTKAISP---R 546
OY 1213 GLTATAVISSP 1224
      ||| : : : : :
Db 547 YSISDRSTISDP 558
```

```

RESULT 13
US-09-801-368-64
: Sequence 64, Application US/09801368
: Patent No. US20020128250A1
: GENERAL INFORMATION:
: APPLICANT: Busby, Robert
: APPLICANT: Call, Brian
: APPLICANT: Hecht, Peter
: APPLICANT: Holtzman, Doug
: APPLICANT: Madden, Kevin
: APPLICANT: Maxon, Mary
: APPLICANT: Milne, Todd
: APPLICANT: No. US20020128250A1man, Thea
: APPLICANT: Royer, John
: APPLICANT: Salama, Sofie
: APPLICANT: Sherman, Amit
: APPLICANT: Silva, Jeff
: APPLICANT: Summers, Eric
: TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
: FILE REFERENCE: 109272.147
: CURRENT APPLICATION NUMBER: US/09/801.368
: CURRENT FILING DATE: 2001-03-07
: PRIOR APPLICATION NUMBER: US 09/487,558
: PRIOR FILING DATE: 2000-01-19
: PRIOR APPLICATION NUMBER: US 60/160,587
: PRIOR FILING DATE: 1999-10-20
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 64
: LENGTH: 1569
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-801-368-64

```

Query Match	3.7%	Score	282.5	DB	10	Length	1589
Best Local Similarity	22.1%	Pred. No.	7e-09				
Matches	133	Conservative	108	Mismatches	244	Indels	117
						Gaps	22

[illegible]

```
Db 1262 SVSAVQVTTKRDNKSPI--HMSSSSLPSSA-SSAFPRLLKLLDIDPYTATQTLVLEH 1318
QY 731 ELFRNIEPTVEYIDDLF--KLRSKTSKANLKRFEVINQETFWVASFLRETNOLKRMKII 788
Db 1319 DLRLRTMECLDRAGTGYCNMGSPNITKFLANANTLTNEVSHITVQADYKTKSKILT 1378
QY 789 KHEIKIALHCRECKNFNSMFAIISGLNLAPVARLRTTWKLPNKYEKLQDLODLEDPGR 848
Db 1379 QYFVYVAQCKELNNSSMTAIVSALYSPIYRLKKTWDLVTSKEDLKNLNLMSDKR 1438
QY 849 NNAKYNNVNSONLQPIIPLFPYIKKDLTFLEHGNDS--KVDGLVNFEXKLMIKAEIR 905
Db 1439 NEVYKRELLRSVT-DVACVPEFEGVYLSDLTFTEVGNPDLELNSTNIINFSKRKIANYI 1497
QY 906 HV 907
Db 1498 EI 1499

RESULT 14
US-09-911-826a-10
; Sequence 10, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; FILE REFERENCE: DDW-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-911-826a-10

Query Match 3.4%; Score 260.5; DB 10; Length 249;
Best Local Similarity 29.9%; Pred. No. 1e-08;
Matches 60; Conservative 48; Mismatches 86; Indels 7; Gaps 4;

QY 710 QISLLQLSTVEVATQLSMRNFELEFRNIEPTVEYIDDLFKLRK-TSCANLKRFEVINQET 768
Db 2 EITLLTLHLELRLAQLTLLEFEMYNKVPSELVGSPTKKDKREVSPNLLKIMKHTTNTV 61
QY 769 FWVASFLRETNOLKRMKIIKHEIKIALHCRECKNFNSMFAIISGLNLAPVARLRTTWK 828
Db 62 RWEKSITEANEVEERLAIMQRAIEVMVWLELNNGILSIYAAMGTASVYRLRWTFQG 121
QY 829 LPNKYEKLQDLODLEDPGRNNAKYNNVNSONLQPIIPLFPYIKKDLTFLEHGNDSKY 888
Db 122 LPRRYKRFLECEKELSDD-HLKKYQERLRSIN--PPCVPEFGRYLTNIHLLEGNDLL 177
QY 889 DG--LVNFEKLRMIKAEIRHV 907
Db 178 ANTELINFSKRKRVAEIIGEI 198

RESULT 15
US-09-911-826a-9
; Sequence 9, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; FILE REFERENCE: DDW-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
```

```
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 9
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-911-826a-9

Query Match 3.3%; Score 254; DB 10; Length 244;
Best Local Similarity 25.9%; Pred. No. 2.4e-08;
Matches 76; Conservative 46; Mismatches 105; Indels 66; Gaps 6;

QY 716 LSTVEVATQLSMRNFELEFRNIEPTVEYI-DDLFKLRSKTSKANLKRFEVINQETFWVASE 774
Db 8 LSAAMELAEQITLLDHIYFRSIPYEEFLGCGMKLDKNERTPIYIMKTSQHFNENSNIYASQ 67
QY 775 ILRETNOLKRMKTIKHEIKIALHCRECKNFNSMFAIISGLNLAPVARLRTTWKLPNKYE 834
Db 68 INNVADISSRPNAIEKWAVADICRLHNNGVLEITSALNRSPIYRLKKTWAKVSKOTK 127
QY 835 KLQDLODLEDPGRNNAKYNNVNSONLQPIIPLFPYIKKDLTFLEHGNDS-KVDGLVN 893
Db 128 ALMDKLOKTYVSSGRRKLNRETL--KNCNPAPVYLGMYLTDLAFLEEGPNTTEGLVN 185
QY 894 FEKLRLMIA--KEIRHYGRMA-SVNMDPALMFTRKKKWSLSLSGSGTNATVLDVAQT 949
Db 186 FSKMRMISHIRIRIQOQAVARIDQO----- 213
QY 950 GGHKKRVRRSSFNLNAKKLYEDAQMARKVQYLSNLELMEDESLSQTLISQCEP 1002
Db 214 -----KVIOYLLDKALVIDEDSLYELSTKIEP 240
```

Search completed: February 18, 2003, 21:40:36
Job time : 303 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: February 18, 2003, 21:22:09 : Search time 40 Seconds
(without alignments)
1102.624 Million cell updates/sec

Title: US-09-911-826A-2

Perfect score: 7721

Sequence: 1 MKPLAIPANHGVGQGEKHS.....PYSGGFSTEDEDEQVSAY 1499

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	354.5	4.6	1596	3	US-09-356-952-3
2	323	4.2	1572	2	US-08-290-731C-5
3	321.5	4.2	1048	2	US-09-356-952-5
4	307.5	4.0	1319	2	US-08-290-731C-2
5	307.5	4.0	1336	2	US-08-290-731C-6
6	303.5	3.9	1333	3	US-09-356-952-2
7	288	3.7	1244	3	US-09-356-952-7
8	285	3.7	1418	2	US-08-290-731C-13
9	282.5	3.7	1589	3	US-09-356-952-4
10	277	3.6	911	3	US-09-356-952-6
11	276.5	3.6	666	1	US-08-318-831-3
12	276.5	3.6	814	1	US-08-318-831-2
13	274	3.5	489	1	US-08-318-831-4
14	273	3.5	402	2	US-08-290-731C-14
15	269	3.5	430	2	US-08-290-731C-9
16	268	3.5	426	2	US-08-290-731C-12
17	265	3.4	364	1	US-08-318-831-6
18	264	3.4	557	4	US-09-356-952-1
19	248	3.2	652	1	US-08-318-831-8
20	245.5	3.2	1297	2	US-08-290-731C-4
21	238.5	3.1	768	5	US-08-408-519-2
22	238.5	3.1	768	5	US-08-408-519-2
23	238.5	3.1	852	2	US-08-408-519-5
24	238.5	3.1	852	2	US-08-408-519-5
25	227	2.9	423	2	US-08-290-731C-10
26	201.5	2.6	2101	1	US-08-195-487-4
27	201.5	2.6	2101	5	PCT-US93-06160-4

28	199.5	2.6	423	2	US-08-290-731C-11	Sequence 11, Appl
29	197.5	2.6	2101	1	US-08-466-390-4	Sequence 4, Appl
30	197.5	2.6	2101	1	US-08-470-950-4	Sequence 4, Appl
31	197.5	2.6	2101	1	US-08-467-781-4	Sequence 4, Appl
32	197.5	2.6	2101	2	US-08-483-924-4	Sequence 4, Appl
33	197.5	2.6	2101	4	US-09-452-294-1	Sequence 1, Appl
34	197.5	2.6	3248	5	US-08-353-700-1	Sequence 1, Appl
35	197.5	2.6	3248	5	PCT-US95-16216-1	Sequence 1, Appl
36	197	2.6	1066	4	US-09-541-782-8	Sequence 8, Appl
37	197	2.6	1066	4	US-09-723-820-8	Sequence 8, Appl
38	194.5	2.5	2482	1	US-08-328-254-6	Sequence 6, Appl
39	191.5	2.5	1184	4	US-09-541-782-2	Sequence 2, Appl
40	191.5	2.5	1184	4	US-09-723-820-2	Sequence 2, Appl
41	186.5	2.4	897	1	US-08-095-737-4	Sequence 4, Appl
42	186.5	2.4	897	1	US-08-480-145-4	Sequence 4, Appl
43	186.5	2.4	897	2	US-08-477-389-4	Sequence 4, Appl
44	186	2.4	1164	4	US-08-923-992A-10	Sequence 10, Appl
45	185.5	2.4	10182	4	US-09-134-001C-3159	Sequence 3159, Ap

ALIGNMENTS

RESULT 1	US-09-356-952-3	Sequence 3, Application US/09356952
Patent No. 6117663		
GENERAL INFORMATION:		
APPLICANT: Boriack-Sjodin, Ann		
APPLICANT: Margarit, S. M.		
APPLICANT: Bor-Sogi, Dafna		
APPLICANT: Cole, Philip		
APPLICANT: Kariyan, John		
TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE		
FILE REFERENCE: 600-1-228N		
CURRENT APPLICATION NUMBER: US/09/356,952		
CURRENT FILING DATE: 1999-07-19		
EARLIER APPLICATION NUMBER: 60/093,631		
EARLIER FILING DATE: 1998-07-21		
NUMBER OF SEQ ID NOS: 14		
SOFTWARE: Patentin Ver. 2.0		
SEQ ID NO 3		
LENGTH: 1596		
TYPE: PRT		
ORGANISM: Drosophila melanogaster		
US-09-356-952-3		
Query Match	4.6%: Score 354.5; DB 3; Length 1596;	
Best Local Similarity	17.1%: Pred. No. 3e-17; Mismatches 474; Indels 823; Gaps 64;	
Matches 309; Conservative 204;		
QY 12 VMGQEKHSPLADFT-----KLHLTSLHPQVTHVSSH-----SGCSITSD 53		
DB 225 MLNQEHLPLPSLPAQASATYEETVELKELHDEKQYORDLMIIRVRELYIVSD 284		
QY 54 SGS-----SLSLDIYATESAGDMD-----LSGLPEPAVSEDDDEEDIE 95		
DB 285 PRELEPIFSNMIMYEVYVTLGSLDVIEMSOEASPCVGSFEEAEFEEDVKKYA 344		
QY 96 RADPPLMSRIVDCEK-----		
DB 345 YDVTQSARDLNNLSKPGASSLTAGHGFRDAVKYKLLVPLCAFYFDYIKHL 404		
QY 114 DPIDRTDDDI---EQLEFHWQLPAFANMTS-----VRRELCAVMAVAYE 157		
DB 405 KDLSSQDDLESEFQVGLLPHLCDEKYMASLSKERQVPSGRYKROL-----AIE 457		
QY 158 RACTIVL-----NDGEELDSWSVILNG-----SVEVTYPDGKAETILGMS 198		
DB 458 RTRELQKVEHWEKDVQGNCFEIRSDLSKSGSKRIWSEKVFLEDC-LMVLCKANT 516		
QY 199 FGVSPT-----MDKEYMKGV---MRTKYVDCC----- 222		

Db 517 KQOPSGATAYDYRLKEXYEMRVDINDRPDSDDLKNSFELAPRMQPIVLTAKNQHK 576
QY 223 -----FVCIADODYRILNOVEKNQK-----VEEGELVAKKEHREL 260
Db 577 HDWADMLWITKSMRLRHLDSIIERKHPRLMRSPREYKFAVPDSCGNIVLEERES- 635
QY 261 DRGTGRKHIVKGT-----ERLTMHLEBHSVVDPTFIEDPLTYRTFLSSPMYEGKK 315
Db 636 -----AGVPMKATLCKLIERLTHI-----YADPTEVKTFLTYRYF-CSFOQLLOL 683
QY 316 LLEMFN--DPSL-----RDKATRY 332
Db 684 LVERFNIPDPSLVQODGTAGAGGCGVGDKEHKNSHREDWKRKEVQPVQFVRLNV 743
QY 333 VLLVNVNHNDEGDPMTATREFEFENNLEREKMGHRLNLNTACAKAKRRLMTLPKS 392
Db 744 LRHVVDHNFYDFEKDPMI--LEKLNFLF-----HVN-----CKSMRK----- 779
QY 393 REAPLPIILGSGSEKGFIVDSVDSKATFAGLKRGDQILEVNGQNFRIQLSKAMEI 452
Db 780 -----WDSY----- 784
QY 453 LRNTHLSIVYKTNLFVFKELLTRLSEKRNAGAPHLKIGDIKKASRYSIPDLAVDEQV 512
Db 785 -----LKI----- 787
QY 513 IGLEKVNKKSKANTVGRNKKLKTLDKTRISILPQKPYNDIGISQSDSIVGLRQTKHI 572
Db 788 --VORKNQOEKSN-----KKTIV-----YATG-----HD 808
QY 573 PTLAPVSGTSSSNPDLLOSHHRLDEFSATPDLPOVLRFVKADQOSRYIMSKDTAKE 632
Db 809 PP--PIEHLSVPR----- 820
QY 633 VVIOAIRFPAVTATPDQYSLCEVSVTEBGVIKQRLPDQLSKLADRIQLSGRYLKNME 692
Db 821 ----- 820
QY 693 TETLCSDEDAQELLRESOISILQSTJEVATQLSMRNPELFERNIEPTXIDDLFKLSK- 751
Db 821 -----DEITLTLHPLELAKQTLLEEFEMKNVPSLESGSPMTKDK 864
QY 752 TSCANLRFEEVINOETFWVASEILRETNOQLRKMKIIFIKIALHGRECKNFEMSPAI 811
Db 865 VKSPNLKIMKHTNVTRWIEKSTTEAENYEERLAIMQRALEVAMVMELENFNGILSIV 924
QY 812 SGLNLAPVALRTWTEKLPNKYEKLFQDLQDLDPDSRNMAKYRVNLQPIPLFP 871
Db 925 AAMGTASYRLRMTFOCLPERYRKFLCECRLSD--HLKYYQERLSIN--PPCVDFFG 980
QY 872 VIKKDLFLHEGNDKSYDG--LVNEFKLMAKELRIHVGRNASVNMPPALMFRTRKKMR 929
Db 981 RYLTNLHLEEGNDLIANTELINFESKRKYAETI----- 1015
QY 930 SLGSLSGSTNATVLDVAQOTGHHKRVRRSSFLLNAKLYE-----DAQMARKVQY 980
Db 1016 --GEIQOYQONPYCLNEST-----IROFEQDLDPFNGLSQKMS--DY 1055
QY 981 LSNLELMEDESLQTLISLQSPATNTLPKNPGK-----KPYKSETPAPAPRAGS 1030
Db 1056 LYNESLRIEPRGCTVYKPPRKMPIPLKSPGKIPRQNDNTSSSKLSNSTSVAAAAA 1115
QY 1031 QOKQOSLPQOQPPRAHKINOGLQVAVSLYPSRKVKPVYKDLPPFGINSQALKITLSL 1090
Db 1116 SSTATST-----ATASAPSLHASSIMDAP----- 1139
QY 1091 SEEGSLERHKKQADDTISNASSQSSPPTSPQSPRKGTYLAP-----SGTYDN 1139
Db 1140 -----TAAANAGSGTLAGEQSPQHNPFAVSFAVPIPERNTSSMSGTPQH 1186
QY 1140 F-SNDSGHSIS-----SNSIYSNSSFDSV-----PVSGLDER 1171

Db 1187 TRTDQNGEVSVPAPHLPKKRGAVHANNNSTLASASAMDVFPALPEHLPPQSLPDSN 1246
QY 1172 RQHSVSIYETNLGMGMERTMIEPDQYSLGYAPMSBGRGLYATATYISSPTEELSQ 1231
Db 1247 PFASTDTEAPPSPL-----PKLVVSP-RHETGNRSP--HGMQNSPHTSTASTVTLT- 1295
QY 1232 DQCDRASLDAADSGRGSWTSKSSGSHDNIQTIOHRSWETLPPG-----HTHPDYSQD 1284
Db 1296 -----GMSTSGGEELCAGGFYFN---SAHQOPGAVPLSPHVVNPMATNMETRAV 1342
QY 1285 PAGL-----WASSSHMDQIMFS-----DHST--KYNQONS 1313
Db 1343 PPPLPRRKERTESCADMAOKROAPDAPLPPRDELSPPPIPRLNHSTGISYLRQSHQ 1402
QY 1314 RESLEQASRASMASSTGYGEDS--EGDGTIKRRGKXSVLEASS-----S 1360
Db 1403 K-----SKEFGNSSLLIPNTSSIMIRNSAIEKRAATASOPNOAAAPTIS 1448
QY 1361 LFSVTEET---KPYMPAHIAVASTTKGLIARKRGYREPPPTP-----PGYIGIP 1410
Db 1449 TTLVTSQAVATDEPLPLPISPAASSTTS-----PLTPAMSPMNPISHP 1496
QY 1411 ITDPBGHSHPAKRPDYVALORSKRVANASDPTAGPSSVQOPHGHPITSRPVKNQOWHK 1470
Db 1497 VESTSSSYAH-----QLRMROOQOQOTHPATYSOHNHATLTP-NHBPQNH 1542
QY 1471 PNESDPRLAP 1480
Db 1543 SNPQSRSSP 1552

RESULT 2
US-08-290-731C-5
; Sequence 5, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENTEEN (MSOS) GENE,
; TITLE OF INVENTION: AND MSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MIOT, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL0921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

LENGTH: 1572 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-290-731C-5

Query Match 4.2%; Score 323; DB 2; Length 1572;
Best Local Similarity 16.9%; Pred. No. 6.4e-15;
Matches 292; Conservative 196; Mismatches 446; Indels 794; Gaps 61;

QY 12 VMQOQKHSLPADFT-----KLHLTDSLHPQVTHVSSH-----SGCSTTSD 53
DB 225 MLMOSEAHILPSPLSLPAQASATYEETVELHDEKQYORDLHMIIRVREELVKIVSD 284
QY 54 SGS-----SSLSDIYQAFSEAGDMD-----LSGLPETAVIDSEDDDEEDIE 95
DB 285 PRELEPFFSINMDIETVTLVLLGLSLEDVIMSDQSAFCVSGCFEELAEFEEDVYKKA 344
QY 96 RASDPLMSRDIIVDCLEK-----113
DB 345 YDVTQSASRDALNNLLSKPGASSLTAGHGFRDAVKYTLPLKLLVPIGAFVYFDYIKHL 404
QY 114 DPLDRDDDI---EQLEEFHQLPAPANNMYS-----VRRELCAMVFAVVE 137
DB 405 KDLSSQDDDIESTFEYOGLHPLHCDLEKVMASLSKERQVPVSGRVFRQL-----AIE 457
QY 158 RAGTIVL-----NDGEELDSWSYLNG-----SVEVYTPDCKAEILCMGNS 198
DB 458 RTELOKQVHEMEDKQVQNCNEPIREDLSLKSQSKRIWSEKRVLELFDG-LMVLCAANT 516
QY 199 FGVSPT-----MDKEYMKGV--MRTKYDDCO-----222
DB 517 KQRTPSGATAYDYRLKEKYFMRVINDIRPSDDLKNSFELAPRMQPIVLAKNAQHK 576
QY 223 -----FVCIADQDYRILANQVEKNQK-----VEEGEITVWKEHREL 260
DB 577 HDWMADLNAVITSMRLHDSILODIERKHPMRSPETLYKFAVPDSCGNIVLEERES- 635
QY 261 DRTGRKGHIVIKGTS-----ERLTMHLYVEHSVDPFTIEDFLTYRTFLSSPMERKK 315
DB 636 -----AGVMTIGATICKLIERLTYHI-----YADPTFVRLTYTRYE-CSFQDLQL 683
QY 316 LLEMFN--DPSL-----RDKVTRV 332
DB 684 LVERFNIPDSLYVQDTGAGAGMGVGDKKHKSHREDWKRYREYVQVQFRFLNV 743
QY 333 VLLAVNNHNFDEGDDPAMTFLEEFENNLREKMGHLLRLIATCAAKARRMLTTPKS 392
DB 744 LRHWVDHNFYDFEKDPMI--LEKLLNLE-----HVN-----GKSMRK-----779
QY 393 REAPLPIILGSEKGCIFVDSVDSGSKATEAGLKRGDQILVNGNGNFENIQLSKAMEI 452
DB 780 -----WVDSV-----784
QY 453 LRNNTHLSITVKTNLVFKELLTRLSEKRNKGAHPILPKIGDIKKASHYIPDLAVDEQV 512
DB 785 -----LKI-----787
QY 513 IGLKYNKKSKKANTVGGRNKLKILDKTRISILPQKPYNDIGIGOSODDSIVGLRQTKH 572
DB 788 --YORKNEOEKSN-----KKIV-----YAYG-----HD 808
QY 573 PTLAVSGTSSSNPDLQSHHRLDPSATPDLRDQVLRVFKADQOSRYIMISKDTAKE 632
DB 809 PP--PIHHSVFN-----820
QY 633 VVIOAIRFAVATAPDOYSLCEVSVTPREGVIKQRRLPDQLSKLADRIOLSGRYLKNME 692
DB 821 -----820
QY 693 TETLCSDEDAQELLRESQISLQSLSTVEVATQLSMRNFELFRNIPEPTYIDDLFKLSK- 751
DB 821 -----DEITLTLHLELAROLDLLEFEMYNKVPSELVGSPTKKDKE 864

QY 752 TSCANLKRFEVINOETFWVASEILRETNOJLKRMIKIKHFIKIALHORECKNFNSMFAIT 811
DB 865 VKSPNLLKIMKHTNTVTRWIERKSITEAENYEBERLAIMQRAIEVMAMLENNNGLISTIV 924
QY 812 SGINLAPVARLRTTWELKLPNKYEKLFODLQDLFDPSSNNMAYRNVLNSQMLQPIIPLFP 871
DB 925 AAGSTASVYLRWTFQGLPERYRKFLRECRSLSD--HLKKYQERLSIN--PPCVDFFG 980
QY 872 VIKKDLTFLHEGNDKVDG--LVNFEKLMIAKEIRIVGMASVNMDDPALMFTRRKKWR 929
DB 981 RYLTNLIHLEEGNPDDLANTLINFSKRRKVAEII-----1015
QY 930 SLGSLSGSTNATVLDVAOTGCHKKRVRRSSFLNAKKLYE-----DAQARKVQY 980
DB 1016 --GEIQOYQNPCLMEEST-----IROFFQDLDPFNGLSKQMS--DY 1055
QY 981 LSNLEMEDESLQTLSECPATNPLPKNPQDK-----KPKSETSPVAPRAGS 1030
DB 1056 LYNESLRIEPRGCKTVKPFKRPKPHILPKSPGILKPRQONTSSSKLSNQTSSVAAAAA 1115
QY 1031 QOKAQSLPQOQPPRAHKNQGLQVPVAVSLYPSRKKVPYKDLPPGINSFQALKITLSL 1090
DB 1116 SSTATSI-----ATASAPSLHASSIMDAP-----1139
QY 1091 SEEGSLERHKKQADRTISNASSQLSPPTSPQSPRKGTYLAP-----SGTVDN 1139
DB 1140 -----TAAANAGSGTLAGEQSPQHNHAFVSFAPVILPERNTSSNGSTPQH 1186
QY 1140 F-SDSGHSEIS-----SRSSIVSNSSPDSV-----PVSLEHDER 1171
DB 1187 TRDQNNGEVSVAPRHLPKKPGAHVMANNNSTLASASAMDVSPALPEHLPOQSLPDSN 1246
QY 1172 RQHRNSIVETNCGMERMERTMIEPDQYLSGYAPMSEGRGLATATVYSSPTELSQ 1231
DB 1247 PRASDTEAPRPSPL-----PKLYVSP--RHETGNRSP--HGRQMSPTSTASTVYLT- 1295
QY 1232 DQGBRASLDAADSGRGSTSCSSGSHNIQTIORSMETLPG-----HNFQVSGD 1284
DB 1296 -----GMSTSGEEECAGGFYN--SAHQGPQAVPILSPHVNVPMATIMETRAV 1342
QY 1285 PAGIMASSSHMDQIMFSDHSTKYRNQNSRESLEQASRASMASSTGWCGEDESGDTGTI 1344
DB 1343 PPLP-----PPRKRERTESCADMAQKQQA-----PDAPTL 1372
QY 1345 KRGGCKVSTLEASSSLTSVTTEBTKVPMRPAHIAVASTTKGLIARKEGRYRPPPTP 1404
DB 1373 PPRDG-----ELSPRPPIPRPL--NHSIGISYLRQSHGSKKE-----1406
QY 1405 GYIGIPITDEPREGSHPARKPPDYNNVALORSRMVARSDD--TAGPSS 1449
DB 1407 -FVGNSSLLPNTSSIMIRR--NSAIEK-RAAAVSQPNQAAAGPIS 1448

RESULT 3
US-09-356-952-5
; Sequence 5, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Borjask-Sjodin, Ann
; APPLICANT: Margalit, S. M.
; APPLICANT: Bor-Sogli, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
; FILE REFERENCE: 600-1-228N
; CURRENT APPLICATION NUMBER: US/09/356,952
; EARLIER FILING DATE: 1999-07-19
; EARLIER APPLICATION NUMBER: 60/093,631
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 5
; LENGTH: 1048
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
US-09-936-952-5

Query Match      4.28; Score 321.5; DB 3; Length 1048;
Best Local Similarity 19.9%; Pred. No. 4.4e-15;
Matches 226; Conservative 166; Mismatches 402; Indels 339; Gaps 45;

13 MGQOEKHSRADPFTKLHUTDS--LHPQVTHVSSHSGCSTITSQSSLSLDYQATESEA 70
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
175 LNHMSHDLRQMKAVAGSAGIAENIDELPASKQCTSSCSSTSHSPAPQRRRGRT 234
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
71 GDMDSLPTAVDSEDDDEEDIERASDPLMSRDIVRCLEKDPIDRTDDIEQLLEFM 130
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
235 IFSVSG-----SSDESLTMSKRKKPYPLNEETL--SLVRARKQLDGKIKOMIKSA 285
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
131 HQLPAPANNMVSARELCAYAVFVVER-AGTIVLNGEELDSMSVTLN-----G 179
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
286 NE---YLSNANFSKMLNFMNFKEYEVSQTPIIDILNLDLTILNLRGLENRVF 342
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
180 SVEVTPDGKAEIL-----CMGNSFGVSPMDREKMGVARTKVDQCFVCIQAO- 229
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
343 DEVAIDDEDEFLKHSLSLSTLSIDYFNK-----KQYFHDVY-----VKETIVAOHL 391
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
230 -----DYCRILNQVEKNMOKVEGEIYVWKEHRELDRTGTRKHGHIYI 272
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
392 TLEDPFVFSPMQNDLPITYEPRMKPSSLINDNAKDK----- 427
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
273 KGTSERLMHLVEHSHVDPFIEDFLTLYRTFLSPMEVQKLLWENDESLDKVTRV 332
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
428 KNGSQNTDIOEEDEDEYEDP---DSLIFPHNLINQDSFND--LKRFN-----LAHV 474
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
333 VLLVNNHFNDECDPAPATRFLEEFENN--LEREKMGSHLLNLIAAARAKRLMTLTK 390
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
475 FKSCGDYFD-----VLKLAIEFVQDLIERE-----NLLVYA---ARMKKNITE 517
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
391 PSREAPLPIILGSEKFGCFIVDSVSGSKATGAGLRGQILLEVNGQNFENQOLSKAM 450
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
518 -----LLLRG-EEGYG---SYDGETAEKS----- 538
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
451 EILRNHTLSITVKTNLVFKEKLTSLSEKRNAGAPHLPIGDIKASRYSPDLAVDE 510
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
539 -----DTNA-VYADSDTKDNEDMRDQVKLP-----RY----- 565
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
511 QVIGLEAKVSKSKANTY-GGRNKKLTKILDKTRISLTPQKPYNDIGIGSQDSIVGLRQT 569
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
566 -----LQREYDESLIWSNNRIKGGSKHALISYLTDNKKDLFF-----DITFLITE 612
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
570 KHIPTALP-VSGTILSSNP-----LLQSHHRLDPSAT----- 602
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
613 RSIPTTFEFLSYLSIQVNLDPEDICFEFEYNEWYTKLIPKCRVLEMTTFEFOYVFLG 672
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
603 PDLPPQVLRVEKADQSRVIMISKDTAKEVVIAIRFEAVTATPDQYSLCEVSTPEGV 662
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
673 YDEPD--LATLNLDFFAQVAIKENTGVSVELLEKNQKF-----KGN 713
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
663 IKQRLPDQLSKLADRIQLSGRYVKKNMET--ETLCSDEDAQELRLRSQSLQSLQSTIVE 720
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
714 IOEATAP-----MKTLDQICODHYSGLTYSYTE--SILAVIDFVL 751
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
721 VATQLSMRNFELFRNIEPEYIDLFK---LRSKTSKANLRFEEVINQOEFVWASELRL 777
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
752 FATQTLIEHEHYCEITTFDCLQKTKWKKYTKSGASPLGEFISFAKKLTWETISYVAK 811
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
778 ETNQLRKMKIHKFTKIALHCRECKNFNSMPAIIISGLNLAPVABRLTTEKLPNKYEKL 837
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
812 EADKRRKRLKLSHFLFIIEYCKKFNFSMTAIIISALVSSPYRLKEKMWAOATPQTRDL 871
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
838 QDLQDLPPSRMAKYRVNLSQNTQPIIPLFPYIKKDLFTFLHGNDKSKVDGLVNFEL 897
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
872 QSLNKLMDPKKNFIVRNELKSLH-SAPCVPFPGYVLSLDTFTDSGNP---DYLVLEHGL 927
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY      898 RMIAKEIRHVGRMASVNMNDPALMFRTRKKMKWRSLSLSQGSTNATVLDVATQGGHKRVR 957
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      928 KGVDHDKKTIIN-----FNKRSR-----LVLDLQ----- 950
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      958 RSSFLNAKKLYEDAQARKVKQYLSN-LELENDESLQTLSDQEPATNLPKNGDKKP 1016
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      951 --ELIEFKTHDFTDRIVIECTISNLENIPRIEKQYQSLIIEP-----KP 996
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      1017 VKSETSPVAPRAGSQOKASLPQPOQPPPAKINKINGLOVPVAVSLPSRKKVP 1069
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      997 RKK-----VVPNSNSNNKSOEKSRDQTDGKTSTKRD-RFPKRLAKTKKKAP 1044
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-08-290-731C-2
; Sequence 2, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENTLESS (MSOS) GENE,
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SUGHRUE, MIYON, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL0921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: O-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-290-731C-2

Query Match      4.08; Score 307.5; DB 2; Length 1319;
Best Local Similarity 19.6%; Pred. No. 6.9e-14;
Matches 224; Conservative 169; Mismatches 409; Indels 339; Gaps 47;

QY      236 NQVEKNMOKVEEGBEIVMWKERE--LDRTGTRKHGHIYIKGTSERLTMHLVEHSHVDP 292
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      240 NDVENIFSRIVDIHLSVKYLGHIEDIVEMTDECSPPHVGSCPEDLA-----ELAFDP 294
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      293 --TFIEDFLN--YRTFLSSPMEVGKKLL-----EMFNDSPLRDKVTRVYVLLVNNHFN 343
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      295 YESARDILRPGFHHGLSLSKPGSAALYLOSIGEGKE-AVQYVLRLLAPVYHCLHY 353
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```



```

OY 344 FEEDPAMTRLEFEENNLEKMGHLL-RLLN-----IACAkakRRLMTLTKPSREA 395
DB 354 FE---LLOKEESEDOEDCKKQATLALLNOSGMEKICSLAKRRL-----400
OY 396 PLPFIILGSEKGFJFVDSVDSKATEAGLRGDOILEVNGONFNIOLSKAM-EILR 454
DB 401 -----SESACRFYSOOM-----KQOLAIKKNMET-OKNIDGMEKIGOCCEPFIM 446
OY 445 NNTHLSTYVK--TNLFVFKLLTRLSEKRNKGAHPILP-----KIGDI 494
DB 447 EGTLTRVGAKHEHRIFLFDGLM--ICCKSNHGQRLPGASSAEYRLKEKPFMKVQINDK 504
OY 495 KKRASRY--STPDLAVDQEOYI-----GLEKVNKKSKANTVGGNKKILDKTRISILPQ 547
DB 505 DQFSEKHAETIILKDONSVYFSKASAEKNNMMAALISIQYRSTLERMLD-----555
OY 548 KPYNDIGIGOSODDSIVGLKQTHIPALPVSGTLSSSNPDLIQSHHRL-DESAHPDLR 606
DB 556 -----VTVLQEEKEBQMRPLPSAEVYRAEPD---SEENILFEENVQPKAG 597
OY 607 DQYLR--VKADQOSRYTISKOTTAKEVYIOAIRFPAVATAPDOYSLCEVSVTEBGVI 663
DB 598 IPITKAGTVLKLRLTYHMYADP-----NEVRFELT---YRSECRPOELSLILI 645
OY 664 KQRLPQOLSKLADRIOLSG-----RYLKNMTEELCS-----D 699
DB 646 ERETEPEPEPTEADRIALENGDOPLSAELKRFREKEYIQVOLVNLVNCRHMEVHFYDFE 705
OY 700 EDAQELRESQ-----710
DB 706 RDADLLQRMEEPIGTVRGKAMKKWESITKIIQKKIARDNGPHNTFPOSSPYEWMI 765
OY 711 -----ISLDLSTVEVATOLSMRNFELFRINIEPEYIDLFKLRSK-TSCANIKRFE 761
DB 766 SRPGHETPFLTLHPLEIRQLTLESIDYRAVOPELGSVWTKEDKINSNPALMKMI 825
OY 762 EVINOETFWVASEILRETNOILKMKIKHFKITALHORECKNNSMFALLISGILAPVAR 821
DB 826 RHHTNLTLMFEKICIVETENLEERVAVSRITELLOVFOELNNNGVLEEVSAANSSSVYR 885
OY 822 LRTTWELPKRYEKLFDODLDFDP--SRMAKYRVNLSONLPILPLFVYIKKDLTF 879
DB 886 LDHTFEQIPSRQKILLEANHESEDHKKYLAKLRSI-----NPRCVPFEGIYLTNLIK 939
OY 880 LHGNDQSKVDGLVNFELKRLIAKEIIRHVGRMASVYNDPALMERTKKKMSLSLSGSGT 939
DB 940 TEEGNP-----EVLRRHKGEL-----INFSEKRRRAVEITGEIOQYON 976
OY 940 NATVLDVAOTGHHKRVRRSSFLNAKKLYEDA-----QMAKKYKOYLSNLELMEDESLQ 994
DB 977 OYPLCLRYEP-----DIKRFENLPMGNSMKETDYLFNSLLE-----1016
OY 995 TLSLOCEPA--TNLTPKNPGDKK--PYKSE--TSPVAFRAGSOOKAASLPPOOQPPPAHK 1049
DB 1017 -----EPHAPKPLPRP--KAYSYPKSPGVRSNRPETMNH-----PPLPLQEP-----1060
OY 1050 INGLQVPAVSLYPSRKVYVYKDLPPPGINSPOALKILSLSEBGLERHKKQAEITISN 1109
DB 1061 -----RKIYSYRIP-----ESETESTASA 1079
OY 1110 ASS---OLSSPTPOSSPPKGYTLAPSGTVDNFSDSGHSEISSRSIVSNSSPDSVPYS 1166
DB 1080 PNBEPRIPLPPASGTSSTNDVCSVPSD---DHSASFHSRSASVSSISLSTKGTDEVPVP 1136
OY 1167 LHDERORHSVSIVETN---LGMGRMERTIMEEDQYSLGSYAPMSBGRGLYATAYISS 1223
DB 1137 PVPFPRRRPSSAPAESPSKIMSKHLDSPALPRPPTSAKAYP-----KTSIDRTISID 1192
OY 1224 P 1224
DB 1193 P 1193

```

```

RESULT 5
US-08-290-731C-6
; Sequence 6, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENTEENLESS (MSOS) GENE,
; NUMBER OF INVENTION: AND MSOS POLYPEPTIDES
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRU, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL0921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: 0-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1336 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-290-731C-6

Query Match 4.08; Score 307.5; DB 2; Length 1336;
Best Local Similarity 19.68; Pred. No. 7e-14;
Matches 224; Conservative 169; Mismatches 409; Indels 339; Gaps 47;

OY 236 NOYENKMKQVEEGEI-VAMKHEHRE--LDRTGTRKGHIVIKGTSERLTMHLEHNSVDP 292
DB 257 NDVENIFSRIVDHELHSLKGLHEDVETDDEGSRHPLVGSCEPDA-----ELALFDP 311
OY 293 --TFIEDFLT--YRFLSSPMEVKGKRL-----EWNDSLBDKYTRVYLLVNNHFN 343
DB 312 YESYARDILRPFHGHPLSQLSKPGALYLOSGECPKE-AVOYVLPRLRLAPVYHCLHY 370
OY 344 FEEDPAMTRLEFEENNLEKMGHLL-RLLN-----IACAkakRRLMTLTKPSREA 395
DB 371 FE---LLOKEESEDOEDCKKQATLALLNOSGMEKICSLAKRRL-----417
OY 396 PLPFIILGSEKGFJFVDSVDSKATEAGLRGDOILEVNGONFNIOLSKAM-EILR 454
DB 418 -----SESACRFYSOOM-----KQOLAIKKNMET-OKNIDGMEKIGOCCEPFIM 463
OY 445 NNTHLSTYVK--TNLFVFKLLTRLSEKRNKGAHPILP-----KIGDI 494
DB 464 EGTLTRVGAKHEHRIFLFDGLM--ICCKSNHGQRLPGASSAEYRLKEKPFMKVQINDK 521

```



```

; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA: P00921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-290-731C-13

```

```

Query Match          3.7%: Score 285; DB 2; Length 418;
Best Local Similarity 27.2%: Pred. No. 5.2e-13;
Matches 91; Conservative 57; Mismatches 131; Indels 56; Gaps 9;

```

```

QY 680 QLSRYYLKNMMEF-----ETLCSDEDAQELLRESQISLQSLSTVEVATQLSMRNFE 731
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 123 EVNCKFKNGEATPAWPKTLDDQICODHYSGLYSTTE-SILAVDPYLFATQTLIEHE 181
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 732 LERNIEPEYIDLFK--LRSKTSCANLKRFEVINOETFWASEILIRETNOIKRMKII 788
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 182 IYCEITTFDCLOKIMKNKYTSYGASPOLNEFISFANKLTNFIYSYKVEADSKSRKRL 241
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 789 KHEFKIALHGRECKNFMFAIISGLNLAPVARLRTWEXLPNKYEKLFODLOLFPDSR 848
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 242 SHEFLIAYECKRNFSSMTDIIISALYSSPIYRLEKTMQAVIPQTRDLQSLNKLMDPKK 301
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 849 NMAKYRVNLSQNLQPIILPFPYIKKDLTFLEHGNDKVDGLVNFELRMIAKEIRHVG 908
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 302 NFIVYRNELKSLH-SAPCVPEFGYVLSDLFTDSGNP---DYLVLEHGLKGVHDEKKYIN 357
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 909 RMASVNDPALMEFRTRKKKRSLSLQSGSTNAIVLDVAQTGGCHKRRVRSSFLNAKKLY 968
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 358 -----FKKRSR-----LVDFLQ-----EIIYFKKTH 378
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 969 EDAQMARKVQYLSN-LELEMDSESLQTLISQCEP 1002
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 379 YDFTKDRVTIECISNSLENIPIHEKQYQSLITEP 413
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 9
US-09-356-952-4
; Sequence 4, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margarit, S. M.
; APPLICANT: Bor-Sogi, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE

```

```

; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-228N
; CURRENT APPLICATION NUMBER: US/09/356,952
; CURRENT FILING DATE: 1999-07-19
; EARLIER APPLICATION NUMBER: 60/093,631
; EARLIER FILING DATE: 1998-07-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1589
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-09-356-952-4

```

```

Query Match          3.7%: Score 282.5; DB 3; Length 1589;
Best Local Similarity 22.1%: Pred. No. 6.6e-12;
Matches 133; Conservative 108; Mismatches 244; Indels 117; Gaps 22;

```

```

QY 333 VLLVNNHFNDEGDPAMTRFL--EEFENLREKMGHLLRLNIAC----- 377
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 988 VLVSDVEFENNL-----FLNNSDDFKDACEK-----YVEISNLACITVDQLIBEREN 1035
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 378 ----AAKAKRRMLTLKPSREAPLPILLGSKGIGF-VSDVDSGSKATEAGLRGQD 432
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1036 LLNYAARMKKNLTAE-----LLKGEQEKWFDIYSEDYSDDDSENDAIIR--DDE 1083
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 433 ILEVNGONFENIQSKAMELIRNNTLSTVKTNLVFKELRLSEKRNCAPIHLKIG 492
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1084 L-----GSEDIYERKANIEKN-----LPWFLTSDYELSIYVDSGKIRGSKTKEA 1128
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 493 DIKASRYSLPDLAVDEVOYIGLEKYNKSKANTVGRNKKILDKTRISILPQK--PY 550
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1129 LIEHLTSHELVDAFNTWLTFRSI-----LTTREFVALI--YRYNLYPREGLSY 1178
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 551 NDIGIGSQDSDSYVLRQTHIPLALPVSGTSSNPDLQSHNRILDSATPDLDPQYL 610
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1179 DVNIMIEKSNPKCKVNVIMKRFLOTYTRNYIEPGI---PLINRAK----- 1225
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 611 RVFKADQSRIMISKDTAKEVQIAIRFAVATAPDQYSLCEVSVTEGVYKQRLPD 670
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1226 -----MYSEKIPADDLQKINCKLINENKE-----PYDP---KQD---D 1261
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 671 QLSKLADRIQSGRYRYKNNMETETLCSDEDAQELLRESQISLQSLSTVEVATQLSMRNF 730
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1262 SVSAVQYQTRDNKSPF--HMSSSLPSSA--SSAFELKRLKLLDDIPYTAVALQVLVLEH 1318
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 731 ELFRNIEPEYIDLFK--KLRSKTSKANLKRFEVINOETFWASEILIRETNOIKRMKII 788
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1319 DLYRITMFECLDRAWQTKYCNMGGSFNITKFLANANTLTNFEVSHITVQADVKTRSKSL 1378
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 789 KHEFKIALHGRECKNFMFAIISGLNLAPVARLRTWEXLPNKYEKLFODLOLFPDSR 848
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1379 QYFVTAHQHCKELNFMSSMAIYSAUSPFIYRLKKTWDVJESNDLKNLNNLMDSKR 1438
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 849 NMAKYRVNLSQNLQPIILPFPYIKKDLTFLEHGNDKVDGLVNFELRMIAKEIR 905
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1439 NFVYRRELLRSVT-DVACVPEFGYVLSDLFTFVGNDFLNSTNIINFSKRTKIANYIE 1497
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 906 HV 907
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1498 EI 1499
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 10
US-09-356-952-6
; Sequence 6, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margarit, S. M.
; APPLICANT: Bor-Sogi, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John

```



```

Db 26 NKIPDEGDT-----PEKPEDPSAL--SKOSEVSNKREESDIDONOSDDGDTETSP----- 74
QY 591 QSHHRLDPSATPDLPDQVLRVKKADQOSRYIMSKD-----TTAKEV-----VIAAIRE 640
Db 75 -----TKSPPTPKSV-----KKNKNSSEPIJFSTINGVMTSCRELDNNRSALSASA 121
QY 641 FAVT-----ATPDQYSLCEVSTPEGVIRKORRLPDQ-----LSKL 675
Db 122 FATATGANGCTPNKEKRYRMSLASAGFPDPQNGDKFEFVIRRAATNRVLNLRHWYSKH 181
QY 676 A-----DRIQLSGRYYLKNMMETETLCSDP--DAQELR-----ESQISLLOL--- 716
Db 182 SQDEFNDELCKVIGTGLEVMHDPPELLTQERKAANIIRTLTQEDPGDNOITLEETQM 241
QY 717 -----STVEAVATQLSMRNPELFRNIEPTETI--DDLKL-----RSKTSKANLKR 759
Db 242 AEGYKAPFPFNHSALEAEOLITLDLHVFKIIEEFGGGMKLEKNEKTPYIMKTTKH 301
QY 760 FEEVINOETFWASIELRETNOILKRMKIIKHFIKIALHCECKNFNSMFAIISGLNAPV 819
Db 302 FNDISN-----LIASEIRNEDINARVSAIEKWAVADICCLNHYNAVLEITSMNRSAI 357
QY 820 ARLTWTEKLPNKYEKIFODLOLDFPSRMAKYRNVLSQNLQPIIPLFPYIKKDLTF 879
Db 358 FRUKTWMKYSKQTKALIDLKQKLVSEGRFKMLREAL--KNCDPPCVPLGYLTDLAF 415
QY 880 LHHGNSDKV--DGLVNEFKLMAKEIRHVGRMASVNMDPALMFRTRKKKRSLSGSLSGS 938
Db 416 IEBCTPYTDEGLVNSKMMISHIIREI----- 444
QY 939 TNATVLDVATGCHKVRSSSEFLNAKLEYDAQOMARKVQYLSNLELMEDESLQTLSTL 998
Db 445 -----KQFOOTAKIRHQA-----KTYOYLLDOSFVWDESLVYSSL 461
QY 999 QCEPATNF 1006
Db 482 RIEPKLPT 489

```

RESULT 14
US-08-290-731C-14
Sequence 14, Application US/08290731C
Patent No. 5843646
GENERAL INFORMATION:
APPLICANT: BOWTELL, David Douglas Lawrence
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
TITLE OF INVENTION: SON OF SEVENTLESS (MSOS) GENE,
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,731C
FILING DATE: 17-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU93/00068
FILING DATE: 17-FEB-1993
PRIOR APPLICATION DATA: P10921/92
FILING DATE: 17-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon

```

; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: 0-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-14

Query Match      3.5%; Score 273; DB 2; Length 402;
Best Local Similarity 28.9%; Pred. No. 3,8e-12;
Matches 80; Conservative 49; Mismatches 122; Indels 26; Gaps 8;

QY 650 YSICEVSVPEGVIRKORRLPDQLSKL-----ADRIQLSGRYYLKNMMETETLCSDPDA 702
Db 102 YKFEVEYVPH--FASAEELHSLSHLHHPSTKRSKMLEGK---ELSQLELDLSLHNSP 157
QY 703 QELLRESQISLLOLSTVEAVATQLSMRNPELFRNIEPTETI----DDL--FKLSKTSKAN 756
Db 158 DPIIYKDELVL-LPPEIRAKQCLIEFGQSFHISRIQFLTKIMDELNFRSPREKSTFY 216
QY 757 LKRFEEVINOETFWASIELRETNOILKRMKIIKHFIKIALHCECKNFNSMFAIISGLN 816
Db 217 LS-----NHLVNEVETIYQEEPRRTNVLAIFYOVCDYLRLELNFAISLISALNS 270
QY 817 APARLTWTEKLPNKYEKIFODLOLDFPSRMAKYRNVLSQNLQPIIPLFPYIKKDL 876
Db 271 SPILRLKRTANLNSKTLASFELLNNLTARKNFSNTRDL--ENCVLPCVPLGYFTD 328
QY 877 LTFLEHGNSDKVDGLVNEFKLMAKEIRHVGRMASV 913
Db 329 LTFLEKTNKNDNFQNMIFDKRTKVTIRLNEIKKFGSV 365

```

RESULT 15
US-08-290-731C-9
Sequence 9, Application US/08290731C
Patent No. 5843646
GENERAL INFORMATION:
APPLICANT: BOWTELL, David Douglas Lawrence
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
TITLE OF INVENTION: SON OF SEVENTLESS (MSOS) GENE,
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,731C
FILING DATE: 17-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU93/00068
FILING DATE: 17-FEB-1993
PRIOR APPLICATION DATA: P10921/92
FILING DATE: 17-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon

GenCore version 5.1.13
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2003, 17:34:25 ; Search time 86 Seconds
(without alignments)
2322.589 Million cell updates/sec

Title: US-09-911-826A-2

Perfect score: 7721

Sequence: 1 MKPLIPANHGWMGQGEKHS.....PYQSGFSTEDEDEQVSAY 1499

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A.Geneseq.101002:*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7721	100.0	1499	21	AAB07792
2	4645	60.2	1675	21	AAB42658
3	4644.5	60.2	1651	23	ABG66725
4	2352.5	30.5	1551	22	ABR64459
5	1508	19.5	286	20	AAH89893
6	1508	19.5	286	22	ABB51059
7	676	8.8	884	21	AAV70964
8	671	8.7	881	21	AAV70965
9	664.5	8.6	849	21	AAV70968
10	637	8.3	957	22	ABB59888

11	545.5	7.1	262	21	AAB07793
12	543	7.0	119	22	AAM59534
13	543	7.0	119	22	AAW72102
14	543	7.0	119	22	AAH32364
15	543	7.0	119	23	ABG41916
16	538.5	7.0	152	21	AAB43070
17	538	7.0	519	22	ABG09372
18	475.5	6.2	261	22	AAE13102
19	475.5	6.2	261	23	AAU99911
20	474.5	6.1	246	23	AAE22105
21	431	5.6	456	22	AAW78706
22	431	5.6	456	22	AAM39061
23	418	5.4	291	21	AAV70967
24	409	5.3	338	21	AAV99651
25	354.5	4.6	1596	21	AAV68821
26	321.5	4.2	1048	21	AAV68823
27	321	4.2	556	21	AAV70966
28	312.5	4.0	167	22	AAO01890
29	310.5	4.0	1336	16	AAH84638
30	309	4.0	1262	23	ABH09768
31	308.5	4.0	227	23	AAU99910
32	307.5	4.0	1319	14	AAH47043
33	303.5	3.9	1333	21	AAH19623
34	303.5	3.9	1333	21	AAV68820
35	299.5	3.9	612	22	ABH71554
36	298	3.9	57	20	AAH89886
37	298	3.9	57	22	ABB51062
38	296.5	3.8	472	22	AAH67823
39	294	3.8	1077	16	AAH91597
40	291	3.8	473	23	ABH04984
41	290	3.8	1275	23	ABH09767
42	288	3.7	1244	21	AAV68825
43	288	3.7	1244	23	ABH09769
44	287.5	3.7	472	23	AAE22102
45	285	3.7	466	22	AAH68566

ALIGNMENTS

RESULT 1	AAH07792	AAH07792 standard; Protein; 1499 AA.
AC	AAH07792:	
DT	07-NOV-2000 (first entry)	
XX	A murine guanine nucleotide releasing factor 4.	
XX	guanine nucleotide releasing factor 4; GRF-4; Ras activator;	
KW	Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;	
KW	cell differentiation; cell transformation; neuronal disorder.	
XX		
OS	Mus sp.	
XX		
EH	Key	Location/Qualifiers
FT	Domain	135..253
FT	Domain	/note="CNP-binding domain"
FT	Domain	266..322
FT	Domain	/note="REM domain"
FT	Domain	386..470
FT	Domain	/note="PDZ domain"
FT	Domain	594..692
FT	Domain	/note="RA domain"
FT	Domain	712..899
FT	Region	/note="CDC25 domain"
FT	Region	1403..1406
FT	Region	/note="py motif"
FT	Region	1425..1428
FT	Region	/note="py motif"
XX		
PN	WO200043510-A2.	

Partial murine qua
Human brain expres
Human bone marrow
Peptide #6401 enco
Human peptide enco
Human ORF2834
Novel human diago
Human phospholip
Human guanine nucl
Human 48921 guanin
Human protein SEQ
Human polypeptide
Rat Ras signalling
Human GTPase assoc
Amino acid sequenc
Human CAMP-GEFI fr
Human polypeptide
MSO1 protein. Mu
Amino acid sequenc
Human 47476 consen
Mammalian son of s
Human Son of seven
Amino acid sequenc
Drosophila melanog
Polypeptide fragme
Human secreted pro
Human guanine-nucl
C3G protein. Homo
Human new ras quan
Amino acid sequenc
Amino acid sequenc
Amino acid sequenc
Human GTP releasin
Human GTP-binding

XX 27-JUL-2000.
PD
XX 20-JAN-2000; 2000MO-CA00042.
PF
XX 20-JAN-1999; 99CA-2259830.
PR
XX (HSCR-) HSC RES & DEV LP.
PA
XX
PI Rotln D, Pham N;
DR WPI; 2000-499228/44.
DR N-PSDB; AAA59383.
XX
PT Nucleic acids encoding guanine nucleotide releasing factor-4 useful for
the treatment of cancers and neuronal disorders -
PS
XX Claim 9; Fig 19a; 89pp; English.
XX
CC The present sequence represents a murine guanine nucleotide releasing
CC factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both
CC in vitro and in vivo. It directly binds cyclic adenosine monophosphate
CC (cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cGMP)
CC binding domain). GRF4 directly connects cAMP-generating (e.g. G protein
CC coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates
CC Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4
CC is a target for Nedda ubiquitination as it binds Nedda. Activation of
CC the Ras signaling pathway controls numerous cellular functions, such as
CC cell metabolism, proliferation, differentiation and transformation.
CC Therefore modulation of Ras activity may provide a mechanism for
CC controlling diseases. GRF4 polynucleotides and polypeptides may be used
CC in the treatment of diseases associated with inappropriate GRF4
CC expression and activity such as cancers and neuronal disorders. The
CC GRF4 polypeptides may be used as antigens in the production of
CC antibodies against GRF4 and in assays to identify modulators (agonists
CC and antagonists) of GRF4 expression and activity. The anti-GRF4
CC antibodies and GRF4 antagonists may also be used to down regulate GRF4
CC expression and activity. Inhibition of Ras can reduce cellulose
CC proliferation and cancers.
XX
XX Sequence 1499 AA:
Query Match 100.0%; Score 7721; DB 21; Length 1499;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKPLAIPANHGVMGOQEHSLPADFTKHLTDSLHPQVTHVSSSHSGCSTISDSSSSLS 60
Db 1 MKPLAIPANHGVMGOQEHSLPADFTKHLTDSLHPQVTHVSSSHSGCSTISDSSSSLS 60
QY 61 DIYATESEAGDMDLSGLPETAVNSEDDDEDIERASDPLMSRDIVADCLEKPIDRTD 120
Db 61 DIYATESEAGDMDLSGLPETAVNSEDDDEDIERASDPLMSRDIVADCLEKPIDRTD 120
QY 121 DDIFQLLEFMHQLPAFANMTSVRRELCAVWFAVVERAGTIVLNDGEELDSWVILNGS 180
Db 121 DDIFQLLEFMHQLPAFANMTSVRRELCAVWFAVVERAGTIVLNDGEELDSWVILNGS 180
QY 181 VEVVYVPGKATILCMGNSFGVSPMDKFKYMKGVKRTKDDCOFCVCIADODCRILINQEK 240
Db 181 VEVVYVPGKATILCMGNSFGVSPMDKFKYMKGVKRTKDDCOFCVCIADODCRILINQEK 240
QY 241 NMQVVEEGELVWYKHEHEDLRTGTRKGIYIKGTSERLTAMHVEHNVPTFIEDLL 300
Db 241 NMQVVEEGELVWYKHEHEDLRTGTRKGIYIKGTSERLTAMHVEHNVPTFIEDLL 300
QY 301 TYRFLSSPMVEGKLLLEWMDPSLRDKVTVRVVLLMVNNHNEEGDPAAMRFLIEEFNN 360
Db 301 TYRFLSSPMVEGKLLLEWMDPSLRDKVTVRVVLLMVNNHNEEGDPAAMRFLIEEFNN 360
QY 361 LERKMGHGLRLNITACAAKAKRRMLTITKSRAPLPFTLLGSEKGFIVDSVDSGS 420
Db 361 LERKMGHGLRLNITACAAKAKRRMLTITKSRAPLPFTLLGSEKGFIVDSVDSGS 420

QY 421 KATEAGLKGDOILEVNGONFENIOLSKAMEILLNNHTLSITVKNLVEKELLTRLSEE 480
Db 421 KATEAGLKGDOILEVNGONFENIOLSKAMEILLNNHTLSITVKNLVEKELLTRLSEE 480
QY 481 KRNGAPHLPKIGDIKASRSIPDLAVDEVOYIGLEVYKNSKANTYVGGRNKLKIIDKT 540
Db 481 KRNGAPHLPKIGDIKASRSIPDLAVDEVOYIGLEVYKNSKANTYVGGRNKLKIIDKT 540
QY 541 RISTLPQKPYNDIGIGOSODSTVGLROTKHIPALPVSGTLSSSNPDLLQSHRRIDFS 600
Db 541 RISTLPQKPYNDIGIGOSODSTVGLROTKHIPALPVSGTLSSSNPDLLQSHRRIDFS 600
QY 601 ATPDLPOQVLKFKADQOSRYIMISKDTAKEVYIQAIREFAVATPDQYSLCEVSTPE 660
Db 601 ATPDLPOQVLKFKADQOSRYIMISKDTAKEVYIQAIREFAVATPDQYSLCEVSTPE 660
QY 661 GVIKORRLPDQSLKADRIQISGRYVYLNMMETETLCSDEDAQLRESQISLQSTVE 720
Db 661 GVIKORRLPDQSLKADRIQISGRYVYLNMMETETLCSDEDAQLRESQISLQSTVE 720
QY 721 VATQLSKRNFELFRNIEPTXYIDLFKRSKTSKANLKRFEVINOETFWVASILLIETN 780
Db 721 VATQLSKRNFELFRNIEPTXYIDLFKRSKTSKANLKRFEVINOETFWVASILLIETN 780
QY 781 QLRKMKIIRKIKTALHCRCKNFNSMFAITSGNLAPVARLRTWBEKLPNKYEKLFDL 840
Db 781 QLRKMKIIRKIKTALHCRCKNFNSMFAITSGNLAPVARLRTWBEKLPNKYEKLFDL 840
QY 841 QDLFDPSSNNAKYRNVLSQNLQPIIPLFPVYIKKDLFLHEGDSYVDGLVNEKRLMI 900
Db 841 QDLFDPSSNNAKYRNVLSQNLQPIIPLFPVYIKKDLFLHEGDSYVDGLVNEKRLMI 900
QY 901 AKETRHVGRMASVMDPALMFRTRKKKWSLGSOSGTNATVYDVOTGGHKKRVRRSS 960
Db 901 AKETRHVGRMASVMDPALMFRTRKKKWSLGSOSGTNATVYDVOTGGHKKRVRRSS 960
QY 961 FLNAKKLYEDAOMARKVQYVLSNLELEMEDESLQTLSCCEPATNTLPKNFGDKRPVKE 1020
Db 961 FLNAKKLYEDAOMARKVQYVLSNLELEMEDESLQTLSCCEPATNTLPKNFGDKRPVKE 1020
QY 1021 TSPVAPRAGSQKQKQSLPQPOQOPPAHKITNOGLQVAVASLYPRKKRVPKYDLPPFGINS 1080
Db 1021 TSPVAPRAGSQKQKQSLPQPOQOPPAHKITNOGLQVAVASLYPRKKRVPKYDLPPFGINS 1080
QY 1081 PQALKITILSEEGSLERHKQKQEDTISMASSQSSPPTSPQSSPRKGYTLAPSGTYDNF 1140
Db 1081 PQALKITILSEEGSLERHKQKQEDTISMASSQSSPPTSPQSSPRKGYTLAPSGTYDNF 1140
QY 1141 SDSGHSSEISRSIVSNSSPDVSFVSLHDERORHVSIVETNLGMGMERRTIMIEPDQY 1200
Db 1141 SDSGHSSEISRSIVSNSSPDVSFVSLHDERORHVSIVETNLGMGMERRTIMIEPDQY 1200
QY 1201 SLGSAFMSSEGRGLYATATVYSSPTEBELSODQDRASLDAADSGRSWTSSCSSGSHDNI 1260
Db 1201 SLGSAFMSSEGRGLYATATVYSSPTEBELSODQDRASLDAADSGRSWTSSCSSGSHDNI 1260
QY 1261 OTIOHORSWETLPPGHFHFQYSGDPAGLMASSIMDOIMFSDHSTKYNROMQRESLEOA 1320
Db 1261 OTIOHORSWETLPPGHFHFQYSGDPAGLMASSIMDOIMFSDHSTKYNROMQRESLEOA 1320
QY 1321 QSRASMASSTGYWGEDSEGDGTIKRRGKQVSYLEAESSLTSTYTEETKVPMPAHIAV 1380
Db 1321 QSRASMASSTGYWGEDSEGDGTIKRRGKQVSYLEAESSLTSTYTEETKVPMPAHIAV 1380
QY 1381 ASSTTKGLIARKEGRYRPPPTPGYIGIPITDFPEGHSHPARKPPDYNAVALQSRNVAR 1440
Db 1381 ASSTTKGLIARKEGRYRPPPTPGYIGIPITDFPEGHSHPARKPPDYNAVALQSRNVAR 1440
QY 1441 SSDTAGESSVOQPGHPTSSRPVKNPQWHPKPNESDPLAPYOSOGFSTEDEDEQVSAV 1499
Db 1441 SSDTAGESSVOQPGHPTSSRPVKNPQWHPKPNESDPLAPYOSOGFSTEDEDEQVSAV 1499

RESULT 2
ID AAB42658 standard; Protein; 1675 AA.
AC AAB42658;
DT 08-FEB-2001 (first entry)
DE Human ORFX ORF2422 polypeptide sequence SEQ ID NO:4844.
XX
Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnerary; antiparkinsonian; antiparkinsonian; nocrotropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antineoplastic;
antiviral; antibacterial; antifungal; antineumatic; antihypoid;
antianemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antiinflammatory disease; coagulation;
thrombosis; contraceptive.
XX
Homo sapiens.
OS
WO200058473-A2.
XX
05-OCT-2000.
PD
31-MAR-2000; 2000WO-US08621.
XX
31-MAR-1999; 9905-0127607.
PR
02-APR-1999; 9905-0127636.
PR
05-APR-1999; 9905-0127728.
XX
30-MAR-2000; 2000US-0540763.
XX
(CURA-) CUBAGEN CORP.
PA
Shinkets RA, Leach M;
PI
MPI: 2000-602362/57.
DR
N-PSDB; AACT6867.
XX
Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
neurodegenerative disorders and cardiovascular disease -
XX
Claim 11: Page 4028-4032; 5507pp; English.
XX
AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytosolic; hepatotropic; vulnerary;
CC antiparkinsonian; antiparkinsonian; nocrotropic; neuroprotective;
CC antiparkinsonian; antiparkinsonian; nocrotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; dermatological; immunosuppressive;
CC antidiabetic; hypotensive; thrombolytic; coagulant; vasotropic;
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC antihypoid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX
SQ Sequence 1675 AA;
Query Match 60.2%; Score 4645; DB 21; Length 1675;
Best Local Similarity 64.3%; Pred. No. 0;
Matches 962; Conservative 163; Mismatches 296; Indels 76; Gaps 26;
QY 20 SLPADFTKLHLPDLSLPOVTHVSSSGSITSDSSSSLSDIYQATESPAGMDLSGLP 79
DB 238 SLPADLTKLHLENPPHPOVTHVSSSGSITASDSSSSLSDIYQATESPAGMDLTSLP 257
QY 80 ETAVDEEDDE-EDIERASDPLMRDIYDCLEKPIRTDDIDRLEFHMQLPAFAN 138
DB 298 EGVDEDEDEDEDEDELDR-TDPLQGRDIVACELEKPEADKTDIDRLEFHMQLPAFAN 356
QY 139 MTMSVRELCAVAVFAVERAGTIVLNDGEELDSWYILNGSEVYYPDGAELLCGNS 198
DB 357 MTMSVRELCSVMIFEVDEAGAILLEDGQELDSWYILNGTIVEISHPDGKVENLEMGNS 416
QY 199 FGVSPTMDREYMGVMRTKVDCCQFVCIAGQDYCRILNOVEKMMQKVEEGELVMYKEHR 258
DB 417 FGITPLDKQYMHGIVRTKVDCCQFVCIAGQDYCRILNOVEKMMQKVEEGELVMYKEHR 476
QY 259 ELDRTRGKHGIVIKGTSERLIMHVEHSVDPFTIEDFLTYRFLSSPMVEYGKLL 318
DB 477 ELDRSGTRKHGIVIKAPPELIMHVEHSIVDPYIEDFLTYRFLSPPLDVGIKLL 536
QY 319 WENDPSLRDKVTRVYLLVANNHFNDEGDPAMTRFLEEFENLREKMGHRLNLNACA 378
DB 537 WFKIDSLRDKVTRVYLLVANNHFNDEGDPAMTRFLEEFENLREKMGHRLNLNACA 596
QY 379 AKAKRRLMTLRKSRREAPLPFILLGSEKGFIVDSVDSKATEGKLRGQIILEVNG 438
DB 597 AKAKRQVYLLQKASRESPLDFSLNGSEKGFIVDSVDSKATEGKLRGQIILEVNG 656
QY 439 QNFENIOLSKAMEILRNNTLSTVKTNLVFVKELTLRLSEKRNCAPIHPIKIDIKKAS 498
DB 657 QNFENITMKAVELLRNNTLSTVKTNLVFVKELTLRLSEKRNCAPIHPIKIDIKKAS 714
QY 499 RYSIPDLAVDEQVIGLEKVNKSKANTV-GGRNKLKLLDKTRISLIPQKPYNDIGIQ 557
DB 715 RHSIOHVPGDIEQT-SQEKSGKKVANTVSGGRNKLKLLDKTRISLIPQKPYNDIGIQ 773
QY 558 SODDSIVGLRQTHIPALVSGTILSSNPDLQSHHRLIDFSGATPDLPOVLRVFAQD 617
DB 774 SODDSIVGTGRCHSLAIMPILPGLTSSSPDLQPTTSMDFSNPSDIPQVLRVFAQD 833
QY 618 QSRYIMISKDTAKKEVYIQAIRFAVATPDQYSLCEVYTPGEGVIRKQRLPQLSKLAD 677
DB 834 QSCYIIISKDTAKKEVYIQAIRFAVATPDQYSLCEVYTPGEGVIRKQRLPQLSKLAD 893
QY 678 RIOLSGRYVYKNNMETLCCSDDAQELRESQISLQISTVEVATQLSMRNLEFENIE 737
DB 894 RIOLNGRYVYKNNMETLCCSDDAQELRESQISLQISTVEVATQLSMRNLEFENIE 953
QY 738 PTEYIDLEFKLRKTSQANLKREFEVINOETFWVASILLFETNOLKMKIKIKHRIKIALH 797
DB 954 PTEYIDLEFKLRKTSQANLKREFEVINOETFWVASILLFETNOLKMKIKIKHRIKIALH 1013
QY 798 CRECKNFNSFAIISGNTLNPVRLRTTWKELPKRYKELFQDLODDLPDPSPRNMAKYNNV 857
DB 1014 CRECKNFNSFAIISGNTLNPVRLRTTWKELPKRYKELFQDLODDLPDPSPRNMAKYNNV 1073
QY 858 NSONLOPPIPLFPVYIKKDLTFLEHNSKVDGLVNEFKLIMAKETIRHGRMASVNDP 917
DB 1074 NSOMQPIPLFPVYIKKDLTFLEHNSKVDGLVNEFKLIMAKETIRHGRMASVNDP 1133
QY 918 ALMFRTRKKKWRSLSGISQSTNATVLDVAQOTGCHKRRVRSSTLNAKKLYEDAQMARKV 977
DB 1134 AMMFROR-----SLSQSTNSNMIDV-QGGAKKKRRARRSSLLNAKKLYEDAQMARKV 1184
QY 978 KOYLSNLELMEDEESLQTLSDCEPATNTLPKNCGDKPKV--SETSPVAPRAGSQQAQ 1035

```
Db 1185 KQYLSLDVETDEEKFQFMSLQMEPAYGTLLTKNLSEKRSAXSSEMSFVPMRSAGQTTKA 1244
QY 1036 SLRPOQGPRAHKINGLOVPAYSLYPRKKYVVKDLPPRGINSPOAKKILSLSEGS 1095
Db 1245 HLHQP-----HKVSOYLOVPANLHPIRKKGQTKD--PALNTSLPQ---KVLSTTEIS 1293
QY 1096 LERHKQAEPTISNASSOLSPSPSPKGYTLAPSGTVDNFSGHSEISRSISIV 1155
Db 1294 ---GKHTEDTISVASSLHSPSPASPGSPHKGYTLIPSAKSDLSDSSHSEISRSISIV 1350
QY 1156 SNSFSDVPVSLHBERORHSVSLVENLGMGRMERRTMEPDQYSLGSIAPMSEG----- 1211
Db 1351 SNCSVDMSALODERSSQALAVPESTGALTEKHAS-----GIGDHSQHGPQWTL 1403
QY 1212 -RGLYATATVISPSTEELSDQGRASLDPAADSGRSGMTSCSSGSHDNITQIHORSWE 1270
Db 1404 KPSLIRKLANVSSSVSNEISQ---EHIIIEADSGRSGMTSCSSSHDNITQISLPKMSWD 1460
QY 1271 TL-PFGHTHEDYSGDPAGLMASSHMDQIMFSHSTYNNQNSRESLEQAQSRASWASS 1329
Db 1461 FLNSYRHTHLD---DPI---AEVEPTDSEPYSCSKSCSRICGQCKGLE---RKSMTSS 1510
QY 1330 TGYWGESESGDTGIRKRGKADVISEASSLSTVTEETKRPVMPAHIAVASTTGGLI 1399
Db 1511 SSL-SDTYEPNYGTVKR-----VLESTPAESSEGLDPKDATDPYKTVTSTTERGLI 1562
QY 1390 A-----RKGRYREPPPTPPGYIGIPITDFPEG-HSHPARKKPPDYVVALQSRHVARSS 1442
Db 1563 VYCVTSRKDDRYREPPPTPPGYIGISLADLKEGPTH--LKPPDYSAVAQRSMHNSL 1620
QY 1443 DTAGPSSVQPHGHPTRSRPVNKPQMHKPNESDRLAPYOSQGTSTEDEDEOVSAV 1499
Db 1621 SRUPPASTLSNLEACVSKIVTQPQRHNLQPFHKPLGDVTD--DSEADENEQVSAV 1675

RESULT 3
ABG66725
ID ABG66725 standard; Protein; 1651 AA.
AC ABG66725;
XX
XX 30-AUG-2002 (first entry)
DT
XX
XX Human novel polypeptide #60.
DE
XX
XX Human: inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; haematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW allergic condition; thrombolytic; thrombosis; coagulation disorder;
KW fungal infection.
XX
XX Homo sapiens.
OS
XX
XX WO200244340-A2.
PN
XX
XX 06-JUN-2002.
PD
XX
XX 30-NOV-2001; 2001WO-US47004.
PF
XX
XX 30-NOV-2000; 2000US-0028952.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Goodrich RM, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI Yamazaki V, Ujwal ML, Drmanac RT;
XX
XX WPI; 2002-508509/54.
DR
XX
XX N-PSDB; ABR94949.
```

```
PT Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing.
XX
XX Claim 10; Page 629-632; 672pp; English.
XX
CC The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic
CC novel polypeptides of the invention.
XX
XX Sequence 1651 AA:
SQ
Query Match 60.2%; Score 4644.5; DB 23; Length 1651;
Best Local Similarity 64.3%; Pred. No. 0;
Matches 962; Conservative 163; Mismatches 296; Indels 75; Gaps 26;
QY 20 SLRADFTKHLTDSLAPQVTHVSSSHSGSITSDSGSSLSLDIYQATESPAGMDLSGLP 79
DB 215 SLRADLTKMLTENPHNPQVTHVSSSGCSIASDSGSSLSLDIYQATESVGVGDVLTRLP 274
QY 80 ETAVDSRDDDE-EDIERASDPLMSRDIYVDCLEKPIDTDDIDQLLEFNMQLPAFAN 138
DB 275 EGPVDSDEDEDEEIDR-TDPLQGRDLVRECLEKPADKTDIDQLLEFNMQLPAFAN 333
QY 139 MTMSVRELCAVWFAVERAGTIVLNDGEELDSWYILNGSEVAVYPODKAETLNGNS 198
DB 334 MTMSVRELCSWIFEVYEDAGATILEDEGLDSWYIILNGYIEISHPGKVNLPMGNS 393
QY 199 FGVSPTMDKRYMKGVMTKVDCCQFYCIAQODYCRILINOVKMKQVEEGETVWVKNR 258
DB 394 FGITPTLDKQYMHGIVATKYVDDQFPCIAQODYRILINHEKNTHHVEEGETVWVKNR 453
QY 259 ELDRGTGRKHIVYIKGSERLTMHIVEBSVVDPTFEDRLTLVYRFLSPMGVKKLE 318
DB 454 ELDRSGTRKHIVYIKATPERLIMHLIEBSIVDPTTYEDRLTLVYRFLSPMDVGITKLE 513
QY 319 WENDPSLRDKVTRVVLVWVNNHFNDEGDPAMTRFLEEFENNLERKMGCHRLNLIACA 378
DB 514 WFKIDSLRDKVTRVTLVWVNNHFNDEGDPAMTRFLEEFENNLERKMGCHRLNLIACA 573
QY 379 AKAKRRLMTLTGRSREAPDLFIILGSEKGFIFVSDVDSGKATAGLKRGDQILEVNG 438
DB 574 AKAKRQVVLQKASRSRPLQFSLNGSEKGFIFEGVEGSRKAASGKLRGDQIMEVNG 633
QY 439 QNFENIOLSKAMELIRNNHLSITVKTNLFVEKELLRLRSEKRNCAPIRLPKGDIKKKS 498
DB 634 QNFENITPKAVELIRNNHLSITVKTNLFVEKELLR-TEQKSSVPHPIKAE-KKSN 691
QY 499 RYSIPDLAVDVEQIVLEKVNKSKANTV-GGRNKLKLLIDKTRISLIPOKPYNDIGIQ 557
DB 692 RHSIQHVPQDIEQT-SQEGSKKVKANTVSGGRNKLIRKILDKTRISLIPKFLPSDGLSQ 750
QY 558 SQDDSTIVGLRQTKHITPALPVSGTILSSNPDLQSHHRIILDSATPDLDPQVLRVFKAQD 617
```

```

Db 751 SODDSIVGTRHCRHSLAIPIGTLSSSSPDLLQPTTSMLDSPNSPDIDQYIRVFKVDQ 810
Qy 618 OSRYIMISDQTAKEVYIOAIRFAVATPDQYSLCEVSEVPEGYIKORLDPDQSLAD 677
Db 811 QSCYIIISDQTAKEVYFAVHEFGITGASDYSICEVSVTEGYIKORLDPDQSKLAD 870
Qy 678 RIQSGRYLYKNNMETETLCSDEDAQELLRESQISLQISTVEVATQLSMRNFELRNIE 737
Db 871 RIQNGRYLYKNNMETETLCSDEDAQELLRESQISLQISTVEVATQLSMRNFELRNIE 930
Qy 738 PREYIDDLFLKLSKTSICALKREVEYINQETFWASSELLRETNOUKRMKIIHFIIKIALH 797
Db 931 PREYIDDLFLKLSKTSICALKREVEYINQETFWASSELLRETNOUKRMKIIHFIIKIALH 990
Qy 798 CRECKNFNSMFAIISGLNAPARLTETWEKLPNKYEKLFQLOLDFEFSRMMAKRVNL 857
Db 991 CRECKNFNSMFAIISGLNAPARLTETWEKLPNKYEKLFQLOLDFEFSRMMAKRVNL 1050
Qy 858 NSQNLQPIPLPFPVYIKDITLHEGNDKVDGLVNFELRMIAKEIRHVGMAVYMDP 917
Db 1051 SSGSQMPPIPLPFPVYIKDITLHEGNDKVDGLVNFELRMIAKEIRHVGMAVYMDP 1110
Qy 918 ALMFTTRKKMRSLSGLSGSTNATVLDVAQTGGRKKRVRRSSFLNAKKLYEDAQMARKV 977
Db 1111 AMMFRR-----SLSGSTNSMMLDV-QCGAHRKRRARRSSLLNAKKLYEDAQMARKV 1161
Qy 978 KQYLSLELEMDSEIQTLSQCEPATNLPKNPGDKKRVK-SEISVPVAPRAGSOQKAOS 1036
Db 1162 KQYLSLELEMDSEIQTLSQCEPATNLPKNPGDKKRVK-SEISVPVAPRAGSOQKAOS 1221
Qy 1037 LPQPOQPPPAHKKINGLVPAVSLYSPRRKRVKDLPRFGINSPPALKKILSLSEESGL 1096
Db 1222 LNPQ-----HRVSYQLVPAVNLHPIRKKQOTKD-PALNTSLRP---KVLGTTEIS- 1269
Qy 1097 ERHKKOEDTISNASSQSLSPTSPOSSPRKGYTAPSGTVNFSDSGHSSEISSRSIYS 1156
Db 1270 --GKHTEDTISVASSLSHSSPPASPOGSPHKGYTLIPSAKSNLSDSSHSSEISSRSIYS 1327
Qy 1157 NSFSQSVPSLDERQRHRSIVETNLMGRMERTMLEPOYSLGSAVAPASEG----- 1211
Db 1328 NCSVDSMSALQDERCSQALNVESTGALERTEHAS-----GIGDHSQHGPQTLLK 1380
Qy 1212 RGLYATATATISSPTEELSDQDGRASLDADSGRGSMTSCSSGSHDNITQTOHQRSMET 1271
Db 1381 PSLINCLAVSSSVSNSEIQQ---EHIIIRADSGRGSMTSCSSSHDNITQTOHQRSMET 1437
Qy 1272 L-PFGHTHFYDYGDPAGLWASSSHMDQIMFSDHSTKYNRQNSRESLEQAQSRASWASST 1330
Db 1438 LMSYRTHLD---DPI---AEVEPTDEPYSCSKSCSRTCGQCKGLE---RKSMTSSS 1487
Qy 1331 GWGDESEBDTGIRKRGKQDVSIEAESSLSVTTEERKPPMPMAHIAVASTTGGILA 1390
Db 1488 SE-SDTYEENYGVKKRR-----VLESTPAESSEGLDKADTPYKYVTSTTEGGLIV 1539
Qy 1391 -----RKEGRYREPPTPPGYIGIPITDPPEG-HSHPAKKPPDYVVALORSMVARRSD 1443
Db 1540 YCVTSKKKDDRIREPPTPGYLGLISLADLKBPETH--LKPEDSVAAVQSRKSMVNSIS 1597
Qy 1444 TAGPSSVQOPHGHPTSSRPVKNKQWPKPNEBPRLAPYOSQGSFSTDEDEDEQVSA 1499
Db 1598 RLPPALSSNLVACVPSKIVTQQRHNLQPFHFKLGDVDA--DSLEADNEQVSAV 1651

```

```

RESULT 4
ABB64459
ID ABB64459 standard; Protein; 1551 AA.
XX
AC ABB64459;
XX
XX 26-MAR-2002 (first entry)
XX
Db Drosophila melanogaster polypeptide seq ID NO 20169.

```

```

XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL08562.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure: SEQ ID NO 20169; 21np + Sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABR2072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_Pct_sequences.
XX
XX Sequence 1551 AA:
SQ
Query Match 30.5%; Score 2352.5; DB 22; Length 1551;
Best local similarity 36.2%; Pred. No. 7.2e-157;
Matches 617; Conservative 222; Mismatches 446; Indels 421; Gaps 52;
Qy 3 PLAIPIANHGVMGQOEKHSLPADFTKLHLTDSLHPQYTHVSSHSGSITSDSGSSLSOI 62
Db 13 PTSIAGTGVVGSSTTINRP-----ELHOKCNR--GSHSPTSSAYSGSDPTMASN 60
Qy 63 YQATESEADMDLSGLPERAVSDEDDDEDFERASDPLMSRDIYVDCLEKPIIDRTDD 122
Db 61 Y-ASSLEAEIDISGLVESVDS---DEEDLAESMDSLTVRDVAVDCLEKPAERSESD 115
Qy 123 IDOLLEFMHQLPAFANPARKTSSVREELCAVWVFAVVERAGTIVLNDGELDSMVSITLGSVE 182
Db 116 VEVLLEFPTGKLAFTNITTAIVRALCSVMFAVVDAGTIVVMSDGEELDSWSYLINGAVE 175
Qy 183 VYTPDGAKIILGNSFGVSPYTWKDEYMGVMTKTKDDQFCVCIADQYCRILINQVEKMM 242
Db 176 IERANGSREELQMGDSFGILPTMDKLYHRGVMTKCDQCFVCIITQTDVYRHOGEEMT 235
Qy 243 QKVEEB-GEIIVWKHRELDRTGT-----RKGIIVIKGTSERTLYMH 283
Db 236 RREHDENGEVVWVTELRISIGAGTDSAGSGSAGSASINMKRGHVIVIRGPERLLOOL 295
Qy 284 VERHSVVDPTFIEDLLTYRTFLSSPMVEGKLLMFN-----DPSLRQVTPRVLL 335
Db 296 VEESMTDPTVYEDLTHRIPTQNPQVTSKLLHWFLEQYDAHKTQELRQDPTRVVLL 355
Qy 336 WVNHNHNDGDPAMTRPLEEFENNLERKMGHLLNLINACAKARRLMTLTKRSRA 395
Db 356 WVNHNHNDGDPAMTRPLEEFENNLERKMGHLLNLINACAKARRMNSCTLTSSRDE 415

```

QY 396 PLPEFILLGSGE-----KGFGLFVDSVDGSKATKTEAGLRGQDILEVNGCNFENIQ 445
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 416 PLNFRIVGVELGVALATGNAVAIVGISHVEFGSKAQDGLKRGQDIIHEVNGQSLDHYT 475
QY 446 LSKAMELRRNTHLSTVKNLFEVFKELLTRLSEB-----KRGAPHLR 489
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 476 SKRALELTGTHTLSISVKNLGFKEIMQALEHGGGTAGSGSISAGSGSKSVRSPRRI 535
QY 490 KIGDIRK-----AS-----KXSIDPLAND----- 508
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 536 CANDIARLHGRSSTTDELSSVSASNRHAWRLSSVMDLLDQPCAPQTPPVSGSGNMA 595
QY 509 ---VEQYIGLEKYN---KSKAN-----TVGQ-----RNKLKILDKTRISILPQK 548
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 596 SNFMQOL--LQSVNNSAKKSGNSNDQODTGGFMTLAPKRRLQKALAKML----- 647
QY 549 PYNDIGIGQSDSIVGL-----RQTKHIFLALPVSG---TLSSSNPDL 589
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 648 -LKNQNHGSSLNDSDTLNDLPKSKLAVSSCSSSTQSSINGCTVSGGRGLYQSQSNPDL 706
QY 590 -----LQSH-HR-----ILDESAT-----PDLPDQVLYEKA 615
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 707 TSLNDGSGDAGNGGGRGLQVNTLMHIIHRPSAATLTSTQSHLLPDVDPHVLKYKA 766
QY 616 DQOSRYIMISKDTAKEVVIOAIRFEAVATPPQYSLCEVSVPEGYIKORRLPDOLSKL 675
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 767 DQCKVLYIKETTAHEVWMLTQIEFGIHDSPSNFSLCEVSVGDGKVKRRRLPDQLONL 826
QY 676 ADRIQUSGRYYLKNMNETFTLQSCDEDAQELLRESQSLQLSTVEVATQLSMRNFELRN 735
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 827 AERISFAARYLKLNDSTEPLVDELALLEYRESNHFILNAYELAIQTLDDAFNRQ 886
QY 736 IETETRIDDFKLRSTSCANLKRFEVINOETFWYASELRETNOLKRMKIIKHFKIA 795
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 887 IETETVYDELFEIERSKXGVPMLSKFAELVNRKFWYSELCAHNIVRRKIKYQKFKIA 946
QY 796 LHCRCCKNFNSMFAIISGLNLAPVRLRTWEKLPNKYEKLFODLQDLFEDPSRNMAYRN 855
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 947 RHCKEGRNFNSMFAIYSGIGHAVSRLRQWEKLPKSYQGLFNDLQDLMDPSRNMASYRQ 1006
QY 856 VLNSQNL-QPITPLFPVYIKKDLTFLEHGDSVYDGLVNEFKRMIAKEIRHGVASVN 914
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1007 LVSALLAOPHPIIPFYPIYKDLTFIHLGNDTVGDLINEKRLMAKEVRLTLHHMCSSP 1066
QY 915 MD-----PALMFTRRKKKMSLSLSOGSTMA---TVLDVQOTGCH---KKVARSSEFLNA 964
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1067 YDLSTLELKGSPSNALESLNOMASQSMAAAGTV--IAANAGQATIKRRKSTAAPNP 1124
QY 965 KKLVEDAOMARKVQYLSNLELEMEDESLQTLSLQCEPA-----TNTLPK--- 1009
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1125 KMFEEQWVRKAYLNLAKILSDDELAKHFLSECEPAHGSYSSISIGNTSHHSQGG 1184
QY 1010 -----NFGDKKPVKSETS-----PVADRAGSQOK 1033
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1185 GSISGAGGSSGGGSSSLNAGDOLSIYHTSSAPNSLSLKRHRPSSPTLSTSS 1244
QY 1034 AOSLPPOOOPRAHKINGLOVPAVSIVSRKKVAVKDLPPRCINSQALKKILSLSE 1093
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1245 TSSTSHQRR---QMHNG-----PKFGTASPOAVKMKMLLSES 1280
QY 1094 GSLERHKKAQEDTISNASSQLSS--PRTSPQSPRKGYTLAPSGTVDNFSDSGHSEISSR 1151
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1281 SKIRPH---QPTVPRHSGTMAGVIRPLHMHMAHGFSPSPGCV-----TSRA 1326
QY 1152 SSIVSNSSSDSVSLHDERROHRSIVETNLGMRMERRMITEPD-----QYSLGSTA 1206
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1327 TSAVANVQCTSPSP-SPCSHRRLASGNNIIPRAIHERSHSDPAPAPPPLPSVLDLSES-- 1383
QY 1207 PMSEGRGLATATVITSPSGTELSQDQGRASLDADSGSGTSSGSHNDIQIOTIQ 1266
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1384 -----SSVIT---FRDLPKRSVTS-----GSISGDSG-----YVHQ 1413

QY 1267 RSMETLFFGHTHFEDYSGDPAGLWASSSHMDQIMFSDHSTKYNRNQONSRESLEQOGRASW 1326
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1414 QZ-----HYHLYQ-----QOQQOQONSQHEPPEPY 1439
QY 1327 ASSTGWGEDSEDDTGTIKRGGKDVSIKAESSLSVTTEETKPYVMP-----AHIAVAS 1382
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1440 TAA-----DCRLDQISNNAVTRNL-NSPCOSTNTPSTPPPPPPQPTATQLSA 1488
QY 1383 STTKGLARKKEGRYRE-----PPPPPPYIGIPIITDFEG--HSHPAKPPDYNV 1430
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1489 PPTAAAYMHARSQQLQOQOQSLAMPPPPPY-NVP---PLGSIYSH----- 1533
QY 1431 ALQSRHVARSSDTAGPSSVQOPGH 1456
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1534 -----HOGTASRHLNMHGH 1549
RESULT 5
AA088983
ID AA088983 standard; protein; 286 AA.
XX
AC AA088983;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polypeptide fragment encoded by gene 140.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; lissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
XX
PN M09854963-A2.
XX
PD 10-DEC-1998.
XX
PF 04-JUN-1998; 98WO-US11422.
XX
PR 18-DEC-1997; 97US-0070923.
PR 06-JUN-1997; 97US-0048877.
PR 06-JUN-1997; 97US-0048881.
PR 06-JUN-1997; 97US-0048884.
PR 06-JUN-1997; 97US-0048893.
PR 06-JUN-1997; 97US-0048896.
PR 06-JUN-1997; 97US-0048899.
PR 06-JUN-1997; 97US-0048915.
PR 06-JUN-1997; 97US-0048949.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048972.
PR 06-JUN-1997; 97US-0049020.
PR 06-JUN-1997; 97US-0049375.
PR 05-SEP-1997; 97US-0057628.
PR 05-SEP-1997; 97US-0057635.
PR 05-SEP-1997; 97US-0057644.
PR 05-SEP-1997; 97US-0057647.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057661.
PR 05-SEP-1997; 97US-0057667.
PR 05-SEP-1997; 97US-0057761.
PR 05-SEP-1997; 97US-0057764.
PR 05-SEP-1997; 97US-0057770.
PR 05-SEP-1997; 97US-0057775.
PR 05-SEP-1997; 97US-0057778.
PR 06-JUN-1997; 97US-0048875.
PR 06-JUN-1997; 97US-0048878.
PR 06-JUN-1997; 97US-0048882.
PR 06-JUN-1997; 97US-0048885.

Query	Match	Best Local Similarity	99.3%	Score 1508:	DB 20:	Length 286:
Matches	284:	Conservative	0:	Mismatches	2:	Indels 0: Gaps 0:
QY	1214	LYAATVATVSSPTEELSDOGDRASLDADSGRGSWTSCSSGSHNITQIOHRSWETLP	1273			
Db	1	LYAATVATVSSPTEELSDOGDRASLDADSGRGSWTSCSSGSHNITQIOHRSWETLP	60			
QY	1274	FGHTHFDYSGDPAGLMASSSHMDQIMFSDHSTKYVRNQSRRESLEBOAQRASWASSTGYW	1333			
Db	61	FGHTHFDYSGDPAGLMASSSHMDQIMFSDHSTKYVRNQSRRESLEBOAQRASWASSTGYW	120			
QY	1334	GEDSEGDGTIKRRGKGVSTLEAESSLITVYTEETKRVYPAPAHIAVASSTTKGLIARKE	1393			
Db	121	GEDSEGDGTIKRRGKGVSTLEAESSLITVYTEETKRVYPAPAHIAVASSTTKGLIARKE	180			
QY	1394	GRYREPPPTPGYIGIPITPDPPEGSHPARKPPDYVNVALQSRMARSPDAPGSSVOOP	1453			
Db	181	GRYREPPPTPGYIGIPITPDPPEGSHPARKPPDYVNVALQSRMARSPDAPGSSVOOP	240			
QY	1454	HGHPTSSRPVANKPQWIKPNESDPRLAPYQSOGFGSTEDEDEDEQVSAV	1499			
Db	241	HGHPTSSRPVANKPQWIKPNESDPRLAPYQSOGFGSTEDEDEDEQVSAV	286			
RESULT 6						
ID	ABB51059	standard; Protein: 286 AA.				
AC	ABB51059:					
XX	07-FEB-2002	(first entry)				
DE	Human secreted protein encoded by gene 140 SEQ ID NO:1012.					
XX						
XX	Human: secreted protein; immunomodulatory; antisclerotic; anti-HIV;					
XX	dermatological; immunosuppressive; antiinflammatory; immunostimulant;					
XX	cytostatic; cardiac; vascular; anti-angiogenic; ophthalmological;					
XX	neuroprotective; noctropic; anticonvulsant; antialzheimers; vulnerary;					
XX	antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;					
XX	multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;					
XX	human immunodeficiency virus; hyperproliferative disorder; wound healing;					
XX	Gaucher's disease; cardiovascular disease; Schmitz syndrome; chemotaxis;					
XX	Chagas' cardiomyopathy; coronary arteriosclerosis; angiolemic disorder;					
XX	corneal graft neovascularisation; diabetic retinopathy; regeneration;					
XX	neurological disorder; Huntington's chorea; Alzheimer's disease;					
XX	Parkinson's disease; infectious disease; chromosome 4.					
OS	Homo sapiens.					
XX						
XX	MO200162891-A2.					
XX						
XX	30-AUG-2001.					
XX						
XX	21-FEB-2001; 2001WO-US05614.					
XX						


```

Db      145  PEPQAGTHDVEEELVEAMALLSQRPDALLT-----VALRKSPGQRTDELDLEELV 199
QY      132  QLPAPAMTMSVARELCAYVAVFVVERAGTIVINDGEELDSVITNGSV- VTYPDGA 190
Db      200  HIKAVHLSVSVARELAAYVLFEPHAKAGVLFSGDKGSMWIIYKGSNNVTRKGGV 259
QY      191  EILCMGNSFVSPTMDKEVYKVMRTKVDCCFVCIADOCYCIILNOVEKNMGKVEEGE 250
Db      260  TTHEDDDGQALVADAPRAATIIIRNNCHFLRYDKODFNRIIDVEAKYMRLEHGK 319
QY      251  IYVYKHEHRELDRTGT-----RKGHIVIKGTSERLTMHVE---EHSVDP--T 293
Db      320  VVLV-----LETSQAGAPSPPTPGRNRYTVMSGTEPKILELLFLAMRDSANDPTEP 374
QY      294  FIDDFLLYRTFLISSPMEGCKLLEWFNDS-----LRDXYRVVLLWY 337
Db      375  FLSDFLTLHSVFPCQLFAALLHHFHVPEPSAGSEQERSTYICNKRQOILRLVSRWY 434
QY      338  NNHFNDEGDPAMTRFLEEPE-----NMLERE---KMGGHLRLNTAC---AAKAK 382
Db      435  ALXSPMLRSPVATSFLOKISDLVSRDTRLNLRQOYPPRRHHNLEN-GCGNVSPTK 493
QY      383  RLMLTLTKPSREARLPRLTLLGGSEKGFIFVDSVDSGSKATEAGLRGQDILEVNGONE 442
Db      494  ARNAPYWFNRHEERLP-----SSAGAIRVGDVY-----521
QY      443  NIDLSAMELRNN-----THLSITYKTNLFVKELLTRLSEKRNKAGRLPKIGDIKKA 497
Db      522  ----PYDICRDPDHSVLTLLHLPVTASV-----REYMAALANE---DHMTK-----558
QY      498  SRSYIPDLAVDVOVIGLEKVNKSKANFVGGNKKLKILDKRISILPKPYNDIGIGQ 557
Db      559  -----GOVL-----VKVNSAG-----569
QY      558  SODDSIVGLKQTKHPIPTALFVSGTSSNPDLQSHRILDFSGATPDLDPQVLRVKADQ 617
Db      570  ----DVYGLQ-----PDARGVATSLGNE-----RIFVVDP 596
QY      618  QSRVIMISKDTTAKEVYIAQIRFANVATIPDQYSLCEVSYTPREGVYIKORRLPQLSKAD 677
Db      597  QEVH-----ELTPH-----PEQLG-----610
QY      678  RIQLSGRYYLKNMNETETLCSDAQELRESQISLLQ-STEVEVATQLSMRNFELEFNI 736
Db      611  -----PTLGSSF-----MLDLVSAKDLQGLTEHDMNLEFNI 642
QY      737  EPTEYIDDLF---KLRSKTSANILKRFEEVINOETFWVASEILRETNLKRMKIIKHFIK 793
Db      643  HOVELLHYVLGPHLDVYT-ANLEFRMRFNELOYVWATLELCVYPPGRADLLKRFIK 701
QY      794  IALHCECKNFNSMFAIISGLNLAPVARLTWTEKLPNKYEKLFQDLQDLFEDPSRNKAKY 853
Db      702  LAHLKEOKMLNSFFAVMEGLSNSAISRLAHATERLPHKRYKLYSALERLEDSMNHRYV 761
QY      854  RNVLNSONLOPPIPLFVYIKKDLTFLEHNGSKVGVLFVEKLMIAKIRIVGRASV 913
Db      762  RLALT--KLSPPVIFPMPDLKDMTTFHESNHTLVELELFEKRMARAVRLHHCORSH 819
QY      914  NMDPALMFTTRKKKRWSSIGLSOGSTNAVLVDAQTGCHKRRVRSFLNAKLLEYDAQM 973
Db      820  STAPLSPLSR-----VSHIHEDSQ 840
QY      974  AR-----KKQYVSNLELDEDSLOTLQCEP 1002
Db      841  SRISTCSEOSLSTRSPASTWAVYOOLKVIDNORELSRLSRELEP 884

```

```

XX      09-AUG-2000 (first entry)
DT
XX      Human Ras signalling pathway associated protein CAMP-GEFI.
DE
XX      Human: Ras signalling pathway: CAMP-GEFI; cyclic adenosine monophosphate;
KW      GEF; guanine nucleotide exchange factor; Rap1A; diagnosis; treatment;
KW      CAMP-GEF-associated disorder; drug; transgenic animal model;
KW      Ras-associated cancer; protein therapy.
XX
OS      Homo sapiens.
XX
XX      Key
FH      Binding-site
FT      231..300
XX
XX      Region
FT      278..282
FT      /note="conserved motif in the CAMP binding pocket;
FT      amino acid at position 280 determines binding
FT      specificity for CAMP/CGMP"
XX
XX      Region
FT      616..639
FT      /label="SCR1
FT      /note="Structurally conserved region which is
FT      highly homologous to Ras-superfamily GEFs"
XX
XX      Region
FT      689..731
FT      /label="SCR2
FT      /note="Structurally conserved region which is
FT      highly homologous to Ras-superfamily GEFs"
XX
XX      Region
FT      768..789
FT      /label="SCR3
FT      /note="Structurally conserved region which is
FT      highly homologous to Ras-superfamily GEFs"
XX
XX      WO200024768-A2.
XX
XX      04-MAY-2000.
PD
XX
XX      22-OCT-1999; 99WO-US24826.
PF
XX
XX      23-OCT-1998; 98US-0105507.
PR
XX      16-NOV-1998; 98US-0108685.
XX
XX      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PA
XX      Kawasaki H, Graybiel A, Housman D;
PI
XX      WPI; 2000-350690/30.
DR
XX      N-PSDB; AAD00315.
DR
XX
XX      Isolated nucleic acid comprises nucleotide sequence encoding protein
PT      selected from normal or mutant CalDAG-guanine nucleotide exchange
XX      factor -
XX
XX      Claim 64: Page 113-115; 128pp; English.
XX
XX      The present sequence is a human cyclic adenosine 3', 5',
XX      monophosphate-guanine nucleotide exchange factor1 (CAMP-GEFI) which
XX      has CAMP binding domain and Ras superfamily GEF
XX      domains. It has substrate specificity for Rap1A and is differentially
XX      distributed in brain and various other tissues. It selectively activates
XX      the Ras superfamily small G protein and Rap1A, functions as a negative
XX      regulator of Ras and directly couples the CAMP signal transduction
XX      system to Ras superfamily cascades. The present sequence
XX      is used for diagnosis and treatment (by protein therapy) of
XX      CAMP-GEF-associated disorders. It is also useful for developing drugs
XX      and producing cell-lines or transgenic animal models for Ras-associated
XX      cancers.
XX
XX      Sequence      881 AA:

```

```

RESULT 8
AA770965
ID      AA770965 standard; Protein; 881 AA.
XX
AC      AA770965;
Query Match      8.7%; Score 671; DB 21; Length 881;
Best Local Similarity 23.5%; Pred. No. 2..7e-38;
Matches 229; Conservative 147; Mismatches 287; Indels 312; Gaps 25;

```



```

Db 129 DDEHEDA-----PLPTREEKKCEDEELQDTMLLSQMGPDAMHMLRKPPGRTVDLLE 183
OY 125 QLEFPHQOLPAFANMTMSVRELCAVMFAVAVRACTIVYLINDEEELDSVILINGSVEVT 184
Db 184 IYIEELLHAKLSLSTYKRELAVLIEFSAHKGGLTVFENOEESTSWIILKGSNNV 243
OY 185 -YPDGKAELICMGNSEFVSPTMDKEYMKGVARTKVDCCQVCAODDYCILNQVEKNMQ 243
Db 244 IYKGYVCTLHEDDFGKLALVNDAPRAASIVLRDNCHFLRVKDEDFNRLDVEANIV 303
OY 244 KVEEGEIVWKEHRELDRGTGRG-----HIVIKGTSERLTMHVE-----EHSVDP 292
Db 304 RLKEHDQDVLEKVPAGNANSGNSQPOQKXTVMGSGPEKLEHLEIIRLEATLNEA 363
OY 293 T--FIEDFLITYRTFLSSPMEVGKLLMFNDLSL-----RDKVTRVVLLWV 337
Db 364 TDSVLNDFIMMHCVFPMFNTQLCPALVAHYHAQPSOGTEOEKMDYALNNKRRVIRLVQWA 423
OY 338 NNHFNDEGDPAMTRFLEEFENLREKMGCHLRLLINACAARRLMTLTK-----PSR 393
Db 424 AMYGDLQEDDVSMARLEEFYVSVD-----ARMT-----AALKQDLPLEKIVKQISE 473
OY 394 EAPLPFILLGSEKGFIFVDSVSGSKATE--AGLKRGDQILEVNGQNFENIQLSKAME 451
Db 474 DAKAP-----QKKHVLQENGTGERAQKQPIRGSEV-----LKVYVC 514
OY 452 ILRNNHLSLTVKTNLFEKLELLRLSEKRNKGNAPHLPIKIGIKASRSIPDLAVDEQ 511
Db 515 MDHTYTTIRVPVATSV---KEVISAADKLGSGE-----545
OY 512 VIGLEKYNKSKANTYGGRRKRLKILDKTRISLPOKPYNDIGISQSDSIYGLQTKH 571
Db 546 -GLIIVKMS-----GGE---KVL-----KP-NDVSVF-----569
OY 572 IPTALPVSGTLSSNPDLQSHHRILDFSATPDLPOGLVRFKADQSRIMISKDTAK 631
Db 570 -TTLTNG-----RLF-----579
OY 632 EAVYQAIREFAVATP-DOYSLCEVSYTPREGVAKORLPQOLSKLADRIOLSGRYIYKNN 650
Db 580 -----ACPREQPD-----SLTP-----LPEQEGPTVGTV---GTFEL---608
OY 691 METETLCSDEDAQELLRESQISLLQSTVEATQLSMRNEFLRNIEPTREYIDLF-KLR 749
Db 609 -----MSSKDLAYQMTYDMLFNCVHELILYHTEGRHN 643
OY 750 SKTSCANLKRFEVINQETFWASSELRETNOLKRMKIIHKFIKIALHCRECKNNSMFA 809
Db 644 FKTTANLDFLRFPNIQFVVYTEICLSQLSKRVQLKFKIKIAHCKEYKNSFEFA 703
OY 810 IISGLNLAPYARLRTWEKLPNKYEKLEFDLQDLFDPSSRMARYRVNLSQNLQPIIPL 869
Db 704 IYVGLSIAVASRALTWKELPSKFKFYAFEFESILMPSRRHRAVR--LTVAKLEPLIIF 761
OY 870 FPIYIKDQFLHGNDSKVGVINEKLEIMIAKEIRHVGMASVNNDPALMFTRRKKWR 929
Db 762 MPLLIKDMFTHGKNTFIDNLVNEFKMRMIANTARTVYRSQFPNP-----809
OY 930 SLGSLGSGSTNATVLDVAQGTGCHKKRVRRSSFLNAKKLYEDAQMARKVOYLNSLELEMD 989
Db 810 -----DAAQ-----ANKNHQ-----VRSYVQLANVIDN 833
OY 990 EESLQTLISLOCEP 1002
Db 834 QRTLSOMSHRLEP 846

```

RESULT 10
 ABB59888
 ID ABB59888 standard; Protein; 957 AA.
 XX
 AC ABB59888;

```

XX 26-MAR-2002 (first entry)
DT XX
DE Drosophila melanogaster polypeptide SEQ ID NO 6456.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
  pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO2001/1042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
DR MPI; 2001-656860/75.
XX
DR N-PSDB; ABL03991.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure: SEQ ID NO 6456; 21pp + Sequence Listing: English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB161737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 957 AA:

```

Query Match 8.3%; Score 637; DB 22; Length 957;
 Best local Similarity 23.8%; Pred. No. 7.8e-36;
 Matches 212; Conservative 156; Mismatches 272; Indels 252; Gaps 29;

```

OY 62 IYATESEAGMDMLSGLEPETAVDSED-DDEEDIERASDPLMSRD---IVRDCLEKDPID 117
Db 244 LYFRIDIEEGSTAAGVPG-----AEDLGANEHIREALSALFORGPATLRMLRKPSHE 299
OY 118 RTDDDEIQLEFPHQOLPAFANMTMSVRELCAVMFAVAVRACTIVYLINDEEELDSVIL 177
Db 300 RTSEELVLEFELVHVALSHLSTSIKRELSTLFEPAHAQACTILFENQDEERSWITL 359
OY 178 NGSVEVT-YPDGKAELICMGNSEFVSPTMDKEYMKGVARTKVDCCQVCAODDYCIIN 236
Db 360 KGSVDVYIHGKGYATLTKGDDGFKLALINDAPRAATIVAKENNHLLRRDKHFNRILR 419
OY 237 QVEKNQKVEEGEELIYWKKEHRELDRGTGRG-----HIVIKGTSERLTMHVEE--H 287
Db 420 DVEANTLRLOEHGKDVLY-----LERYAKORQOHSAFKYTVMGTFPAKMLEHLETRIGQ 474
OY 288 SY--VDPTEFELLYRRTFL-----SSPMEVGKLLMFNPSLRDKYTRVLLWVNN 339
Db 475 SVGMDP-FLDDFLTLHYIMPVYQDLDAQTPEDREYIT-----NEKKRVIOEQMKVMA 527
OY 340 HFNDFEGDPAMTRFLEEFENLREKMGCHLRLLINACAARRLMTLTKPSREAPLPF 399
Db 528 VRHAAFEEPSVCFIDILAAVEADP-----553

```

```

OY 400 ILGSEKGEIIVDSVDSGSKATEAGLRGDDILEVNGONFENIOLSKAMEILRNTHL 459
D 554 -----DLNEBTSTVHN----- 564
OY 460 SITVKNLFEVKELTFRLS--EKKRNGAPHLPIGDIKASRSYSPDLAVDEQVIGLEK 517
D 565 -----VLTMARQEDRN-----QN 579
OY 518 VNKSKANTYGRNKKLKILDKTRISILPKPKYNDIGSGSDSDIVGLRQTHIPALP 577
D 580 AGQKWK-----LP-----PNCOP 592
OY 578 V---SGTSSNPDLLOSHHRIIDFSATPDLDPQVLVFKADQOSRYIMSKDTAKEV 634
D 593 ICLFSGNATPSKTVI-----RPD-DDITFVYCADHTYCTLRPMTTAELEIK 639
OY 635 IQATREFAVTATPPQVSLCEVSTPE-GVIKOR--RLPDOLSKLADRIOLSGRY--LKN 689
D 640 ACADKRLQNLNGPDLVLYVEKNGERSVFKDNVSIPTGIS-----LNGRLFVSKD 692
OY 690 NMETETLCSDDAQELLRES-QISLLOSTIVEVATQLSMRNFELEFRNIEPTETIDLP-K 747
D 693 HLDALTLQGE---OECPTGEVDILEITLSTKELAYHITLFEMWDFWAVEHELYHTEGR 749
OY 748 LRSKTSANLKRFEVINOETFWVASIELRETNOLKRMKIIKHFIKIALHORECKNFSM 807
D 750 HHFGKITANLQVFLRRENEVOYVITELVSTPSLSKRGVLKRAIKLAAYCKEYONLMAF 809
OY 808 FAISGLMLAFVARTLTWTKLPKYEKLFODLOLDFPSHMAKYRNVLNSQNLQPPII 867
D 810 FAVVWGSNNMVASRLQGWKEXIKPSKFKRIFOEFALIDPSNRHRAVYFVG--KLQPLI 867
OY 868 PLFPIYKDLTFLHEGNSKVDGLVNFELKMLAKETRHVG--KMASVNMMP 917
D 868 PFMPLLLKDMTFAHGKNKTSLDGLVNFEMKMMMAQMTMTIRFCRSRLGEP 919

RESULT 11
ID AAB07793 standard; Protein: 262 AA.
AC AAB07793;
DT 07-NOV-2000 (first entry)
DE Partial murine guanine nucleotide releasing factor 4.
XX guanine nucleotide releasing factor 4; GRF-4; Ras activator;
KW Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;
KW cell differentiation; cell transformation; neuronal disorder.
XX Mus sp.
OS WO200043510-A2.
PN 27-JUL-2000.
PD 20-JAN-2000; 2000WO-CA00042.
PF 20-JAN-1999; 99CA-2259830.
PR 20-JAN-1999; 99CA-2259830.
PA (HSCR-) HSC RES & DEV LP.
XX Rotin D, Pham N;
XX WPI; 2000-499228/44.
DR N-PSDB; AAA59384.
XX Nucleic acids encoding guanine nucleotide releasing factor-4 useful for
PT the treatment of cancers and neuronal disorders -
PS Claim 34; Fig 19b; 89pp; English.

```

```

XX AAB07793-95 represent partial murine guanine nucleotide releasing
CC factor (GRF)-4 (Ras activator) polypeptides, encoded by the 3
CC reading frames of clone 7.7 (AA59384). GRF4 activates Ras both
CC in vitro and in vivo. It directly binds cyclic adenosine monophosphate
CC (cAMP) directly via its CNMP-BD (cAMP/guanine monophosphate (cGMP)
CC binding domain). GRF4 directly connects cAMP-generating (e.g. G protein
CC coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates
CC Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4
CC is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of
CC the Ras signalling pathway controls numerous cellular functions, such as
CC cell metabolism, proliferation, differentiation and transformation.
CC Therefore modulation of Ras activity may provide a mechanism for
CC controlling diseases. GRF4 polynucleotides and polypeptides may be used
CC in the treatment of diseases associated with inappropriate GRF4
CC expression and activity such as cancers and neuronal disorders. The
CC GRF4 polypeptides may be used as antigens in the production of
CC antibodies against GRF4 and in assays to identify modulators (agonists
CC and antagonists) of GRF4 expression and activity. The anti-GRF4
CC antibodies and GRF4 antagonists may also be used to down regulate GRF4
CC expression and activity. Inhibition of Ras can reduce cellulose
CC proliferation and cancers.
XX
SQ Sequence 262 AA;
Query Match 7.1%; Score 545.5; DB 21; Length 262;
Best Local Similarity 66.9%; Pred. No. 2.9e-30;
Matches 107; Conservative 13; Mismatches 37; Indels 3; Gaps 2;
OY 1325 SWASSTGWCDESGDTGTRKRGKDVSTIEASSLSTVTTEKRPVMPAHIAVAST 1384
D 6 SMSSTAAVAAALELVDPDPCGCHNSGCKDVSAEASSSMVPTTEAKPVMPAHIAVTPST 65
OY 1385 TKGLIAKKEGRYREPPETPGYIGITPTEPEGHSHPARKPPYVALORSRVARSST 1444
D 66 TKGLIAKKEGRYREPPETPGYIGITPTEPEGHSHPARKPPYVALORSRVARSST 125
OY 1445 AGPSSVQOPHGHTSSRNPNKPMQHKPNESDPRILAYQSO 1484
D 126 PABG--QTPPA-AAASRPSGSKPMQHKPSDADPRILAFQPO 162

RESULT 12
ID AAM59534 standard; Protein: 119 AA.
AC AAM59534;
DT 05-NOV-2001 (first entry)
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31639.
XX Human brain expressed exon: gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX Homo sapiens.
OS WO200157275-A2.
PN 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00667.
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA

```

```

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
PS Example 4; SEQ ID NO: 31639; 650pp + Sequence listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 119 AA;

```

```

Query Match 7.0%; Score 543; DB 22; Length 119;
Best Local Similarity 85.7%; Pred. No. 1.2e-30;
Matches 102; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 685 YLKNMTEFTLCSDDAQELRESQISLQSTVEVATQLSMRNPELFRNIPTETID 744
    |||||
DB 1 YLKNMTEFTLCSDDAQELRESQISLQSTVEVATQLSMRNPELFRNIPTETID 60
QY 745 LFLRSKTSKANLKRREEVNQTFFWVASEILRETQNLKMKIKIKFIKIALHCRECKN 803
    |||||
DB 61 LFLRSKTSKANLKRREEVNQTFFWVASEILRETQNLKMKIKIKFIKIALHCRECKN 119

```

RESULT 13

AA072102
ID AAM72102 standard; Protein: 119 AA.

AC AAM72102;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32408.

KW Human: bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukemia; lymphoma; myeloma.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human bone marrow -

```

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 119 AA;

```

```

Query Match 7.0%; Score 543; DB 22; Length 119;
Best Local Similarity 85.7%; Pred. No. 1.2e-30;
Matches 102; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 685 YLKNMTEFTLCSDDAQELRESQISLQSTVEVATQLSMRNPELFRNIPTETID 744
    |||||
DB 1 YLKNMTEFTLCSDDAQELRESQISLQSTVEVATQLSMRNPELFRNIPTETID 60
QY 745 LFLRSKTSKANLKRREEVNQTFFWVASEILRETQNLKMKIKIKFIKIALHCRECKN 803
    |||||
DB 61 LFLRSKTSKANLKRREEVNQTFFWVASEILRETQNLKMKIKIKFIKIALHCRECKN 119

```

RESULT 14

AA032364
ID AAM32364 standard; Protein: 119 AA.

AC AAM32364;

DT 17-OCT-2001 (first entry)

DE Peptide #6401 encoded by probe for measuring placental gene expression.

KW Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488997/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human placenta -

XX Claim 27; SEQ ID NO 32633; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs;

XX see AAI31315-AA157546). The present sequence is a peptide encoded by one

XX CC such probe. The probes are useful for producing a microarray for

XX CC predicting, measuring and displaying gene expression in samples derived

XX CC from human placenta. The probes are useful for antenatal diagnosis of

XX CC human genetic disorders.

XX SQ Sequence 119 AA;

XX Query Match 7.0%; Score 543; DB 22; Length 119;

XX Best Local Similarity 85.7%; Pred. No. 1.2e-30;

Matches 102; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

OY 685 YLLKNMTEFLCSDEDAQELRLRSQISLQSLSTVEVATQLSMRNPELFNIEPTETIDD 744
 Db 1 YLLKNMTEFLCSDEDAQELRLRSQISLQSLSTVEVATQLSMRNPELFNIEPTETIDD 60

OY 745 LFKLRKSTCANLKRFEFVINOETFWVASEILRETNO LKRMKIIKHFIKIALHCRECKN 803
 Db 61 LFKLRKSTGNTHLKRFEDIYNOETFWVASEILRETNO LKRMKIIKHFIKIALHCRECKN 119

RESULT 15
 ABG41916
 ID ABG41916 standard; Peptide: 119 AA.

AC ABG41916;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 31581.
 XX

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 OS
 XX Homo sapiens.
 XX
 PN NO200186003-A2.
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001MO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2002-114183/15.
 PT
 PT measure gene expression in human lung samples -
 XX
 PS
 PS Claim 27: SEQ ID No 31581; 634pp: English.
 XX

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,

CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX

SEQ Sequence 119 AA;

Query Match 7.0%; Score 543; DB 23; Length 119;
 Best Local Similarity 85.7%; Pred. No. 1.2e-30;
 Matches 102; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

OY 685 YLLKNMTEFLCSDEDAQELRLRSQISLQSLSTVEVATQLSMRNPELFNIEPTETIDD 744
 Db 1 YLLKNMTEFLCSDEDAQELRLRSQISLQSLSTVEVATQLSMRNPELFNIEPTETIDD 60

OY 745 LFKLRKSTCANLKRFEFVINOETFWVASEILRETNO LKRMKIIKHFIKIALHCRECKN 803
 Db 61 LFKLRKSTGNTHLKRFEDIYNOETFWVASEILRETNO LKRMKIIKHFIKIALHCRECKN 119

Search completed: February 18, 2003, 21:23:10
 Job time : 96 secs


```
Db 647 TTGTAATTTAGGGTTTGTAGACCTTTTGTATTTATTTATTTAAAGAAACATTTATAGT 706
QY 5607 GGATACATTTGCAAGTAAGCAAGCTTGGATGTGTGAGCTAAATGCCAGCTTTATACG 5666
Db 707 GGATACATTTGCAAGTAAGCAAGCTTGGATGTGTGAGCTAAATGCCAGCTTTATACG 766
QY 5667 CTTCTTCAAGACAGCCCTCTTATTTGAAATTTGCAATTTAGGGAATTAACAGCCCTTAAC 5726
Db 767 CTTCTTCAAGACAGCCCTCTTATTTGAAATTTGCAATTTAGGGAATTTAGGGAATTTAGG 817
QY 5727 GTGATTAAGATCAAAACCTGTGTAGACATGCCAGCTTTTGCAGGAGGTTAGTACCC 5786
Db 818 ATGATTTAAAG-TCACAAACATGGTTA-ATACGCCAGCTTTGCAAA--TGGTTAGTACCC 872
QY 5787 AAGACTAACCTCCAGTGGCTTTATGAGCGCTGCATATAGAGAGGCTTAAGTATAGCA 5846
Db 873 AAGGCTGACCTGTAGTGGCTTTGGGAGCGCATATGAGAGAGGCC-AAGTGTAGCG 931
QY 5847 ACCATCTGCTACAGCTGCTATTTACCTATANTAGCTG-AAATGACCCCTCAGCTCTAT 5905
Db 932 CCGGTCTGCTACATAGCTGCTGCTAGCCCTTAGATGACTGAAATGACCCCGGCTCTAT 991
QY 5906 TTTTGTGTG-TTTTGCACAGACTCCGGAAGAGTGAAGGCTGCCAATCTGATAGTACTC 5964
Db 992 TCTTGTGTGTGTTTGCACAGACTCCGGAAGAGTGAAGGCTGCCAATCCGATAGTACTC 1051
QY 5965 AATGTGAGAGTCTGCTGCTGTGATTTTTTCCATTAAATTCAGCTGATATATGA 6024
Db 1052 AGATGTGAGGCGCTGCTGATCTTGA--TTTTCATTTAAATTCAGCTGATATATGA 1108
QY 6025 TCAGTATTAAGCTAAATAGCTTCAAAATTTTAAAGTGAATTTGAGAGTGTGTTTTCAC 6084
Db 1109 TCGATGATTAAGCTAAATAGCTTCAAAATTTTAAAGTGAATTTGAGAGTGTGTTTTCAC 1168
QY 6085 GATTCAAACAGTCTAGCTGCTTTATTTAATTTCTTCTGTATGATGAGCATTTGCTA 6144
Db 1169 GTGTCAACAGTCTAGCTGCTTTATTTAATTTCTTCTGTATGATGAGCATTTGCTA 1228
QY 6145 CTGCTATTTACATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTG 6204
Db 1229 CTGCTATTTACATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTG 1287
QY 6205 CAGTTTATTTATATAGCTATGAGCTCATGTCATATAGAAAGACAAATCTAGCTCT 6264
Db 1288 CAGTTTATTTATATAGCTATGAGCTCATGTCATATAGAAAGACAAATCTAGCTCT 1347
QY 6265 ACCACAAGTTGCACAAATGTTATCTAAGCATTAAGTAATTTAGAAACATAGACTGCTA 6324
Db 1348 ACCACAAGTTGCACAAATGTTATCTAAGCATTAAGTAATTTAGAAACATAGACTGCTA 1407
QY 6325 TCTCAGTTGCTGTGTGATGTCAGAGTCAGATGTACATTTAAGCTGTATTTCCATCA 6384
Db 1408 TCTCAGTTGCTGTGTGATGTCAGAGTCAGATGTACATTTAAGCTGTATTTCCATCA 1467
QY 6385 CTTTATATATCTTATGATGTCATGTCATTTTGAAGAGACATTTGGAGTCTGATATCC 6444
Db 1468 CTTTATATATCTTATGATGTCATGTCATTTTGAAGAGACATTTGGAGTCTGATATCC 1527
QY 6445 TTTTGTATTTTATACAAATATTTGATACATATTTGTTATATTTTGTGAAGATGGTGA 6504
Db 1528 TTTTGTATTTTATACAAATATTTGATACATATTTGTTATATTTTGTGAAGATGGTGA 1587
QY 6505 AATGTACTATGTTATGCTTCTACATCCAGTTTGTACAGCTGGAATAATTAATATATA 6564
Db 1588 AATGTACTATGTTATGCTTCTACATCCAGTTTGTACAGCTGGAATAATTAATATATA 1647
QY 6565 ACAT 6568
Db 1648 ACAT 1651
```

RESULT 2
AL583554

```
LOCUS AL583554 789 bp mRNA linear EST 16-FEB-2001
DEFINITION AL583554 LFI_FL013_Fbrn1 Homo sapiens cDNA clone CS0DF006YF09 5
prime, mRNA sequence.
ACCESSION AL583554
VERSION AL583554.1 GI:12952629
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 789)
AUTHORS Li,W.-B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF006YF09"
/clone_lib="LFI_FL013_Fbrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by life
technologies. Contact : Reng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
liang@lifetechn.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 247 a 130 c 134 g 273 t 5 others
ORIGIN
Query Match 11.6%; Score 760.8; DB 9; Length 789;
Best Local Similarity 99.3%; Pred. No. 1,8e-143;
Matches 759; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 5805 GGCCTTATGAGCCCTGCATATAGAGAGCCCTAAGTGTAGCAACCATCTGTACAGCTG 5864
Db 1 GGCCTTATGAGCCCTGCATATAGAGAGCCCTAAGTGTAGCAACCATCTGTACAGCTG 60
QY 5865 CTATTAACCTATTAAGACGAATGACCCCTCAGCTATTTTGTGTGTTTGTGACA 5924
Db 61 CTATTAACCTATTAAGACGAATGACCCCTCAGCTATTTTGTGTGTTTGTGACA 120
QY 5925 GACTCGGAAGATGGAAGCTGCCAATCTGAGTACTGCAATTAAGTGAAGAGTGTGCT 5984
Db 121 GACTCGGAAGATGGAAGCTGCCAATCTGAGTACTGCAATTAAGTGAAGAGTGTGCT 180
QY 5985 CTTGCAATTTTTCATTAATTCAGCTGATATGATGATGATGATGATGATGATGATGATG 6044
Db 181 CTTGCAATTTTTCATTAATTCAGCTGATATGATGATGATGATGATGATGATGATGATG 240
QY 6045 GCTTCAATTTTAAAGTGAATTTGCAAGTCTTTTTCAGTATTCGAACAAATGTCACTGC 6104
Db 241 GCTTCAATTTTAAAGTGAATTTGCAAGTCTTTTTCAGTATTCGAACAAATGTCACTGC 300
QY 6105 TTTATTTAATATTTCTCTGTATCATGATGATGATGATGATGATGATGATGATGATG 6164
Db 301 TTTATTTAATATTTCTCTGTATCATGATGATGATGATGATGATGATGATGATGATG 360
QY 6165 TTATGCAATTTGTAATTTTATGATGATGATGATGATGATGATGATGATGATGATG 6224
Db 361 TTATGCAATTTGTAATTTTATGATGATGATGATGATGATGATGATGATGATGATG 420
QY 6225 TGGACCTCATGCTCATATAGAAAGACGAATATAGCTCTACCAACAAGTTGCACAAATGT 6284
```

QY	2491	TCATCAGTGGCC	TAACCTGG	GCACCC	CAGTGG	CAGACG	CTGG	GAAC	CGACTG	GGG	GAATTC	2550
Db	1	TCATCAGTGGCC	TAACCTGG	GCACCC	CAGTGG	CAGACG	CTGG	GAAC	CGACTG	GGG	GAATTC	60

```

FEATURES
  source
    Location/Qualifiers
      1..743
        /organism="Homo sapiens"
        /db_xref="taxon:9606"

```

```

/clone="CSDF006YF09"
/clone.lib="LTI_FLO13_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/notes="Organ: Fetal brain; Vector: PCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
PCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
Maryland 20850, USA Fax : (1) 301 610 8371 Email :
f.liang@lifestech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT      271 a      132 c      125 g      209 t      6 others
ORIGIN

Query Match      11.1%; Score 727.4; DB 9; Length 743;
Best Local Similarity 99.1%; Pred. No. 1e-136;
Matches 736; Conservative 5; Mismatches 1; Indels 1; Gaps 1;

OY 5805 GGCCTTATGAGCGCTGCATATAGAGAGCGCTTAGTACACCAATCTGCTCACAGCTG 5864
|||||
DB 743 GGCCTTATGAGCGCTGCATATAGAGAGCGCTTAGTACACCAATCTGCTCACAGCTG 684
|||||

OY 5865 CTAATTAACCTTAATAGATGAGAAATGACCCCTCCTATTTTGTGTTGTTGCACA 5924
|||||
DB 683 CTAATTAACCTTAATAGATGAGAAATGACCCCTCCTATTTTGTGTTGTTGCACA 624
|||||

OY 5925 GACTCCGGAAGAGTGAAGCGCTGCATCTGAGTAGTCTCAATGAGAGAGTGCCTGT 5984
|||||
DB 623 GACTCCGGAAGAGTGAAGCGCTGCATCTGAGTAGTCTCAATGAGAGAGTGCCTGT 564
|||||

OY 5985 CTGCGATTTTTCCTATTAATTCAGCTGATCATATGATGCTGATTAACGTAATA 6044
|||||
DB 563 CTGCGATTTTTCCTATTAATTCAGCTGATCATATGATGCTGATTAACGTAATA 504
|||||

OY 6045 GCTTCAATTTTAAAGTGAAGTGAAGTGTTCCTACGTTATCAACCAAGTCAAGTGC 6104
|||||
DB 503 GCTTCAATTTTAAAGTGAAGTGAAGTGTTCCTACGTTATCAACCAAGTCAAGTGC 444
|||||

OY 6105 TTATATTAATTAATTCCTCTGATCATGAGCATTTGCTACTGCTTATTAATGTCAA 6164
|||||
DB 443 TTATATTAATTAATTCCTCTGATCATGAGCATTTGCTACTGCTTATTAATGTCAA 384
|||||

OY 6165 TTATGCAATTTGTAATTTTACATGTAATGCAATTTTGGCAGTTTATTAATAGGCTA 6224
|||||
DB 383 TTATGCAATTTGTAATTTTACATGTAATGCAATTTTGGCAGTTTATTAATAGGCTA 324
|||||

OY 6225 TGGACCTCATGTCATATAGAGAGCAAGAAATCTAGCTCCACCAAGTGGACAATGT 6284
|||||
DB 323 TGGACCTCATGTCATATAGAGAGCAAGAAATCTAGCTCCACCAAGTGGACAATGT 264
|||||

OY 6285 TATCTAAGCATTAAGTAATTTAGAACATAGAGCTGTAATCTCAGTTCGCTGTCATG 6344
|||||
DB 263 TATCTAAGCATTAAGTAATTTAGAACATAGAGCTGTAATCTCAGTTCGCTGTCATG 204
|||||

OY 6345 TCAAGTGCAGAAATGTACAAATTAACGTGTCATTTCCCTCATATCTTTGATPACTTGTACC 6404
|||||
DB 203 TCAAGTGCAGAAATGTACAAATTAACGTGTCATTTCCCTCATATCTTTGATPACTTGTACC 144
|||||

OY 6405 TGAATGCTTTTGAAGAGACATGAGAGTGCATGATCCCTTTGTAATTTTATACAAAT 6464
|||||
DB 143 TGAATGCTTTTGAAGAGACATGAGAGTGCATGATCCCTTTGTAATTTTATACAAAT 84
|||||

OY 6465 AATTTGACATATTTGTTATTTTGTGGAAGATGAGAAATGTAC-TATGTTTATGCT 6523
|||||
DB 83 AATTTGACATATTTGTTATTTTGTGGAAGATGAGAAATGTACATGTTTATGCT 24
|||||

OY 6524 TCTACATCCAGTTTGTACAAAGCT 6546
|||||
DB 23 TCTACATCCAGTTTGTACAAAGCT 1
|||||

```

```

RESULT 5
LOCUS      BG403618      812 bp      mRNA      linear      EST 12-MAR-2001
DEFINITION 602419334F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4526168 5',
ACCESSION  BG403618
VERSION     BG403618.1  GI:13297066
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 812)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LLM10433 row: d column: 09
            High quality sequence stop: 687.
            Location/Qualifiers
                1. 812
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_image="4526168"
                /clone_lib="NIH_MGC_93"
                /issue_type="translational cell papilloma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: bladder; Vector: pCMV-Sport6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.7 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH-MGC Library."

BASE COUNT      248 a      139 c      153 g      272 t

Query Match      10.8%; Score 708.8; DB 12; Length 812;
Best Local Similarity 95.9%; Pred. No. 5.9e-133;
Matches 771; Conservative 0; Mismatches 27; Indels 6; Gaps 4;

OY 5758 GCGAGCCTTGCAGAGGAGTTGATGACCAAGAAAGTCAACCTCCAGAGGCTTATGACG 5817
|||||
DB 1 GCGAGCCTTGCAGAGGAGTTGATGACCAAGAAAGTCAACCTCCAGAGGCTTATGACG 60
|||||

OY 5818 CTCATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGCTTTAAACCTAT 5877
|||||
DB 61 CTCATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGCTTTAAACCTAT 120
|||||

OY 5878 AATGACTGAATGACCCCTCCTACTATTTTGTGTTGTTTGCACAGACTCCGAAAG 5937
|||||
DB 121 AATGACTGAATGACCCCTCCTACTATTTTGTGTTGTTTGCACAGACTCCGAAAG 180
|||||

OY 5938 TGAAGCTGCATCTGAGTAGTACACCAATGTGAGAACTGCTGCTGGAATTTT 5997
|||||
DB 181 TGAAGCTGCATCTGAGTAGTACACCAATGTGAGAACTGCTGCTGGAATTTT 240
|||||

OY 5998 TCCATTAAATTCAGCTGATATTGATGATGATGATGATGATGATGATGATGATGAT 6057
|||||
DB 241 TCCATTAAATTCAGCTGATATTGATGATGATGATGATGATGATGATGATGATGAT 300
|||||

OY 6058 AAAGTGAATTCAGTGTGTTTTCAGTGTATCAACCAATGCTGCTTATTTATTAAT 6117
|||||
DB 301 AAAGTGAATTCAGTGTGTTTTCAGTGTATCAACCAATGCTGCTTATTTATTAAT 360
|||||

```


[illegible]

QY	6269	CAAGTTGGCAAAAGTGTATCTTACAGATTAAGTAATTTAGAAACATGAGACGTGTAATGTC	6328
Db	324	CAAGTTGGCAAAAGTGTATCTTACAGATTAAGTAATTTAGAAACATGAGACGTGTAATGTC	265
QY	6329	AGTTGCGCTGTGATGATCGTAAGTGCAGAAATGTACAAATTAACAGTGATTTCCGTACTCTT	6388
Db	264	AGTTGCGCTGTGATGATCGTAAGTGCAGAAATGTACAAATTAACAGTGATTTCCGTACTCTT	205
QY	6389	TGATACACTGCTGTACCTGTATGCTCTTTTAGAAGACATTTGGTGCAGTCTGTATCCCTTTT	6448
Db	204	TGATACACTGCTGTACCTGTATGCTCTTTTAGAAGACATTTGGTGCAGTCTGTATCCCTTTT	145
QY	6449	GTAATTTTAATTCACAAATTAATTTGATCATTTGGTATATTTTGGTGAACATGGTGAATTTG	6508
Db	144	GTAATTTTAATTCACAAATTAATTTGATCATTTGGTATATTTTGGTGAACATGGTGAATTTG	85
QY	6509	TACTATGTTTATGCTTCTACATCATCGTTTGTACAAAGCTGGAATAATTAATATAACAT	6568
Db	84	TACTATGTTTATGCTTCTACATCATCGTTTGTACAAAGCTGGAATAATTAATATAACAT	25
RESULT 8			
BM71124/c		697 bp	linear
LOCUS			EST 28-FEB-2002
DEFINITION		UI-E-EJ0-shk-a-24-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone	
ACCESSION		BM71124	
VERSION		BM71124.1	GI:19030382
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (bases 1 to 697) Bonaldo,M.F., Lennon,G. and Soares,M.B.	
TITLE		Normalization and subtraction: two approaches to facilitate gene discovery	
JOURNAL		Genome Res. 6 (9), 791-806 (1996)	
MEDLINE		97044477	
COMMENT		Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@iuii.wiwi.edu Tissue Procurement: Dr. Gregg Hageman cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: M13 Reverse. Location/Qualifiers 1..697 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="UI-E-EJ0-shk-a-24-0-UI" /clone_lib="UI-E-EJ0" /tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid" /dev_stage="fetal and adult" /lab_host="DH10B (Life Technologies) (T1 phage resistant)" /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: Ecor I; Site_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT73-Pac	

vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dt)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTACGGA; eye anterior segment, AATGCGGCAT; optic nerve, CCATTAGTGT; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).

BASE COUNT 167 a 153 c 150 g 227 t
ORIGIN

Query Match 10.5%; Score 688; DB 14; Length 697;
Best Local Similarity 100.0%; Pred. No. 9,8e-129; Indels 0; Gaps 0;

Matches 688; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 2384 AATTTCAGAGAAACAAACAGCTGAAGAGATGATCATTTAAGCATTTTCACAGAT 2443
      |||
Db 688 AATTTCAGAGAAACAAACAGCTGAAGAGATGATCATTTAAGCATTTTCACAGAT 629

QY 2444 AGCAGTGCATGTAGGAGATGCAAGAAATTTTAACTCAATGTTTGCATATCAGTGCCT 2503
      |||
Db 628 AGCAGTGCATGTAGGAGATGCAAGAAATTTTAACTCAATGTTTGCATATCAGTGCCT 569

QY 2504 AAACCTGGACCAAGTGGCAGACCTGGGAGAGAACTTCCCATTAATACGA 2563
      |||
Db 568 AAACCTGGACCAAGTGGCAGACCTGGGAGAGAACTTCCCATTAATACGA 509

QY 2564 AAAACATTTCAAGATCTCCAGACCTGTTTGAATCTTCCAGAAATGSCAAATATCG 2623
      |||
Db 508 AAAACATTTCAAGATCTCCAGACCTGTTTGAATCTTCCAGAAATGSCAAATATCG 449

QY 2624 TAATGTTCTCAATATGTAATAATCTACAACTCCCATTAATGCTTATCCAGATTAACA 2683
      |||
Db 448 TAATGTTCTCAATATGTAATAATCTACAACTCCCATTAATGCTTATCCAGATTAACA 389

QY 2684 AAAGATCTCACTCTCTCAGAGAAATGATCAAAAGTATGACGCGGTGTCATTT 2743
      |||
Db 388 AAAGATCTCACTCTCTCAGAGAAATGATCAAAAGTATGACGCGGTGTCATTT 329

QY 2744 TGAGAGCTAAGATGATGCAAAAGAAATTCGTCACGTTGGCCGAATGGCTTAGTGAA 2803
      |||
Db 328 TGAGAGCTAAGATGATGCAAAAGAAATTCGTCACGTTGGCCGAATGGCTTAGTGAA 269

QY 2804 CATGAGACCTGCGCTCATGTTTCAGAGCTCGGAGAAAGATGGCGGATTTGGGCTCCT 2863
      |||
Db 268 CATGAGACCTGCGCTCATGTTTCAGAGCTCGGAGAAAGATGGCGGATTTGGGCTCCT 209

QY 2864 CAGCAGAGGTAGTACAATGACACAGTGTAGATGTTGCTCAGACAGGTGCTCATAAAA 2923
      |||
Db 208 CAGCAGAGGTAGTACAATGACACAGTGTAGATGTTGCTCAGACAGGTGCTCATAAAA 149

QY 2924 GCGGATGAGTGTAGTGTCTTCTCAATGCCAAAAGCTTTATGAAGTGGCCAAATGGC 2993
      |||
Db 148 GCGGATGAGTGTAGTGTCTTCTCAATGCCAAAAGCTTTATGAAGTGGCCAAATGGC 89

QY 2984 TCGAAAAGTGAAGAGTACCTTTTCCAAATTTGAGCTAGAAATGACGAGAGACTCTTCA 3043
      |||
Db 88 TCGAAAAGTGAAGAGTACCTTTTCCAAATTTGAGCTAGAAATGACGAGAGACTCTTCA 29

QY 3044 GACATTTATCTCTGAGTGTGAGCCAGCA 3071
      |||
Db 28 GACATTTATCTCTGAGTGTGAGCCAGCA 1
```

RESULT 9
BM423354 1064 bp mRNA linear EST 29-JAN-2002
LOCUS BM423354
DEFINITION AGNCOURT 6402148 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5516489
5', mRNA sequence.
ACCESSION BM423354
VERSION BM423354.1 GI:18391579
KEYWORDS EST.

SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1064)
NIH-MGC <http://mhc.nci.nih.gov/>.
AUTHORS
TITLE
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
plate: LNCM2017 row: c column: 18
High quality sequence stop: 799.
Location/Qualifiers
FEATURES
source
1..1064
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_41"
/lssue_type="amelanotic melanoma, cell line"
/lab_host="PH10B (Phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library"

BASE COUNT 278 a 256 c 272 g 244 t 14 others
ORIGIN

Query Match 10.4%; Score 681.4; DB 13; Length 1064;
Best Local Similarity 97.1%; Pred. No. 2e-127;
Matches 725; Conservative 0; Mismatches 18; Indels 4; Gaps 3;

```
QY 3439 GGAAGGCTATACCTTGGCTCCAGAGTGTCTGTGATATTTTTCAGATTCGTGTCACA 3498
      |||
Db 180 GGAAGGCTATACCTTGGCTCCAGAGTGTCTGTGATATTTTTCAGATTCGTGTCACA 229

QY 3499 GTGAAATTTTTCACGATTCAGATTTGTTAGCAATGCTTTTGAAGTGCAGAGTCT 3558
      |||
Db 240 GTGAAATTTTTCACGATTCAGATTTGTTAGCAATGCTTTTGAAGTGCAGAGTCT 299

QY 3559 CACTGCAGATGAGAGGCGCAGAGGCAATTCGTACAGATCGTGAACAACCTTAGGGA 3618
      |||
Db 300 CACTGCAGATGAGAGGCGCAGAGGCAATTCGTACAGATCGTGAACAACCTTAGGGA 359

QY 3619 TGCGCAGATGAGAGGCGGACATGATTAAGCTATAGTATAGTTGGGCTCTATG 3678
      |||
Db 360 TGCGCAGATGAGAGGCGGACATGATTAAGCTATAGTATAGTTGGGCTCTATG 419

QY 3679 GTGGCAGATGAGAGGCGGACATGATTAAGCTATAGTATAGTTGGGCTCTATG 3738
      |||
Db 420 GTGGCAGATGAGAGGCGGACATGATTAAGCTATAGTATAGTTGGGCTCTATG 479

QY 3739 CAGAGAACTTTCCAGAGATCAGGGGATCGCGGTGATGCTGATGCTGACAGTGGCC 3798
      |||
Db 480 CAGAGAACTTTCCAGAGATCAGGGGATCGCGGTGATGCTGATGCTGACAGTGGCC 539

QY 3799 GTGGAGCTGAGAGTCAATGCTCAAGTGGCTCCCATGATTAATACAGACATCCAGACC 3858
      |||
Db 540 GTGGAGCTGAGAGTCAATGCTCAAGTGGCTCCCATGATTAATACAGACATCCAGACC 599

QY 3859 AGAGAAGCTGGAGAGCTTCCATTCGCGCATCTTCACTTGTGATTTATTCAGGGAGTCTG 3918
      |||
Db 600 AGAGAAGCTGGAGAGCTTCCATTCGCGCATCTTCACTTGTGATTTATTCAGGGAGTCTG 659
```


AUTHORS Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
 TITLE EST (Duesterhoeft, et al.)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Duesterhoeft A

MIPS
 Am Kiofererplatz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.
 No sl sequence available.
 This clone (DKFZp564l142) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, Germany; Email: clone@rzpd.de.
 Location/Qualifiers

FEATURES

source
 1..702
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp564l142"
 /clone_id="564 (synonym: hibr2)"
 /tissue_type="brain"
 /dev_stage="fetal"
 /lab_host="X1-2blue"
 /note="Vector: PAMPI; Site_1: NotI; Site_2: SalI"
 BASE COUNT 194 a 124 c 147 g 237 t
 ORIGIN

Query Match 9.9%; Score 648; DB 9; Length 702;
 Best Local Similarity 98.4%; Pred. No. 1.2e-120;
 Matches 665; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 5453 GAGCGAGAGCGGTGCATTTGGCGATGTCTCCTCCATTGAGATGGCAAAACCC 5512
 1 GGGCGAGAGCGGTGCATTTGGCGATGTCTCCTCCATTGAGATGGCAAAACCC 60
 DB 1 ATTTTAACTATATTTCTTTGTTTGTATTTAGAGTGTGCTTTGTTTGT 5512
 QY 5513 ATTTTAACTATATTTCTTTGTTTGTATTTAGAGTGTGCTTTGTTTGT 5512
 61 ATTTTAACTATATTTCTTTGTTTGTATTTAGAGTGTGCTTTGTTTGT 120
 DB 121 TTTTGTTTTTTTTAAAGAAACATTTTAACTGATGCAATGCAAGAGCAGCTT 180
 QY 5573 TTTTGTTTTTTTTAAAGAAACATTTTAACTGATGCAATGCAAGAGCAGCTT 5632
 121 TTTTGTTTTTTTTAAAGAAACATTTTAACTGATGCAATGCAAGAGCAGCTT 180
 DB 121 TTTTGTTTTTTTTAAAGAAACATTTTAACTGATGCAATGCAAGAGCAGCTT 180
 QY 5633 GGGATGTGGAGCTAATGCCAGCTGTTTACTGCTCTTTCAAGACAGCTCCCTTTAT 5692
 181 GGGATGTGGAGCTAATGCCAGCTGTTTACTGCTCTTTCAAGACAGCTCCCTTTAT 240
 DB 181 GGGATGTGGAGCTAATGCCAGCTGTTTACTGCTCTTTCAAGACAGCTCCCTTTAT 240
 QY 5693 GAATTTGGCATTAGGGAATAACAAGCCTTTAAACGTGATAAAGATCAAAAACCTGGTTA 5752
 241 GAATTTGGCATTAGGGAATAACAAGCCTTTAAACGTGATAAAGATCAAAAACCTGGTTA 300
 DB 241 GAATTTGGCATTAGGGAATAACAAGCCTTTAAACGTGATAAAGATCAAAAACCTGGTTA 300
 QY 5753 GACATGGCAGCCTTTGCAAGAGGTTAGTACCAAGACTAAGTCAAGTGGCTTTAT 5812
 301 GACATGGCAGCCTTTGCAAGAGGTTAGTACCAAGACTAAGTCAAGTGGCTTTAT 360
 DB 301 GACATGGCAGCCTTTGCAAGAGGTTAGTACCAAGACTAAGTCAAGTGGCTTTAT 360
 QY 5813 GGAGCGCATATAGAGAGGCTTAAGTGAACACATCTGCTCACAGCTGCTATTAA 5872
 361 GGAGCGCATATAGAGAGGCTTAAGTGAACACATCTGCTCACAGCTGCTATTAA 420
 DB 361 GGAGCGCATATAGAGAGGCTTAAGTGAACACATCTGCTCACAGCTGCTATTAA 420
 QY 5873 CCTATATGACTGAATGAGACCCCTCCACCTATTTTGTGTTTGGACAGACCTCCG 5932
 421 CCTATATGACTGAATGAGACCCCTCCACCTATTTTGTGTTTGGACAGACCTCCG 480
 DB 421 CCTATATGACTGAATGAGACCCCTCCACCTATTTTGTGTTTGGACAGACCTCCG 480
 QY 5933 AAAAGTGAAGCTGCCATCTGAGTACTCAATGTGAAGAACTGCTGCTGGATT 5992
 481 AAAAGTGAAGCTGCCATCTGAGTACTCAATGTGAAGAACTGCTGCTGGATT 540
 DB 481 AAAAGTGAAGCTGCCATCTGAGTACTCAATGTGAAGAACTGCTGCTGGATT 540
 QY 5993 TTTTTCATTAAATTCAGTGCATATTTGATCAGTGAATGAAGCTGAATAGCTTCAA 6052
 541 TTTTTCATTAAATTCAGTGCATATTTGATCAGTGAATGAAGCTGAATAGCTTCAA 600

QY 6053 TTTTAAAGTGAATTCAGTGTCTTTCAGTGTATCAAAACATGTCAGCTTATTTA 6112
 1 TTTTAAAGTGAATTCAGTGTCTTTCAGTGTATCAAAACATGTCAGCTTATTTA 659
 DB 601 -TTTAAAGTGAATTCAGTGTCTTTCAGTGTATCAAAACATGTCAGCTTATTTA 659
 QY 6113 ATAAATCTCTCTGTA 6128
 1 ATAAATCTCTCTGTA 675
 DB 660 ATAAATCTCTCTGTA 675

RESULT 12
 BM838008 646 bp mRNA linear EST 06-MAR-2002
 K-EST0114179 S95NU601 Homo sapiens cDNA clone S95NU601-74-All 5',
 LOCUS
 DEFINITION
 mRNA sequence.
 ACCESSION
 BM838008.1 GI:19194417
 VERSION
 BM838008.1
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 646)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 JOURNAL
 Unpublished (2002)
 COMMENT
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@email.krdb.re.kr
 Plate: 74 row: A column: 11
 High quality sequence stop: 646.
 Location/Qualifiers

FEATURES

source
 1..646
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="S95NU601-74-All"
 /clone_id="S95NU601"
 /sex="M"
 /tissue_type="ascites"
 /cell_line="Epithelial"
 /lab_host="SNU-601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI; Site_2: XhoI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including SfiI site by treatment of 74 RNA ligase and the first strand cDNA was synthesized with Superscript II using SfiI oligo-dT primer. After first strand synthesis, RNA was degraded by NaOH treatment and cDNA was amplified by PCR reaction. The PCR products were digested with SfiI and cloned into DraIII- digested pME18S-FL3 vector. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 220 a 104 c 144 g 178 t
 ORIGIN
 Query Match 9.8%; Score 646; DB 14; Length 646;
 Best Local Similarity 100.0%; Pred. No. 3.1e-120;
 Matches 646; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 903 ATGCATTGTTGGTGAAGAGCATTCAGTACTAGATCCAAATTCATAGAGAACTTCTGTG 962
 1 ATGCATTGTTGGTGAAGAGCATTCAGTACTAGATCCAAATTCATAGAGAACTTCTGTG 60

QY	1963	ACCTATGAGACTTTTCTTTCTAGCCCAATGGAAAGGGCAAAAAGTTATTTGGAGGGTTT	1022
Db	61	ACCTATGAGACTTTTCTTTCTAGCCCAATGGAAAGGGCAAAAAGTTATTTGGAGGGTTT	120
QY	1023	AATGACCCGAGCCTCAGGGATTAAGGTTACACGGGTAGTATTATTGTGGTAAATATATAC	1082
Db	121	AATGACCCGAGCCTCAGGGATTAAGGTTACACGGGTAGTATTATTGTGGTAAATATATAC	180
QY	1083	TTCAATGACTTTGAAGGAGATCTTCGAATGACTCGATTTTATAGAAGATTTTGAACAACAT	1142
Db	181	TTCAATGACTTTGAAGGAGATCTTCGAATGACTCGATTTTATAGAAGATTTTGAACAACAT	240
QY	1143	CTGGAAAGAGAAATATGGGTGACACCTAAGGCTGTGGAATATACGCTGTGCTCTTAA	1202
Db	241	CTGGAAAGAGAAATATGGGTGACACCTAAGGCTGTGGAATATACGCTGTGCTCTTAA	300
QY	1203	GCAAAAAGAAATTTGATGATCTTAAACAAACCATCCGAGAAAGCTCTTTGCCCTTTATC	1262
Db	301	GCAAAAAGAAATTTGATGATCTTAAACAAACCATCCGAGAAAGCTCTTTGCCCTTTATC	360
QY	1263	TTACTTGGAGGCTCTGAGAAAGGATTTGGAATCTTTGTGACAGTATGATTCAGGTAGC	1322
Db	361	TTACTTGGAGGCTCTGAGAAAGGATTTGGAATCTTTGTGACAGTATGATTCAGGTAGC	420
QY	1323	AAAGCACTGAAAGCAGGCTTGAAACGGGGGATCAGATATTAGAAGTAAATGGCCAAAC	1382
Db	421	AAAGCACTGAAAGCAGGCTTGAAACGGGGGATCAGATATTAGAAGTAAATGGCCAAAC	480
QY	1383	TTTGGAAACATTTTCAGCTCTCAAAAGCTATGGAATTTCTAATAATACACATTTTACT	1442
Db	481	TTTGGAAACATTTTCAGCTCTCAAAAGCTATGGAATTTCTAATAATACACATTTTACT	540
QY	1443	ATCACTGTGAAACCAATTTATTTGTATTTTAAAGAACTTCTAACAAGATTTGCAAGAG	1502
Db	541	ATCACTGTGAAACCAATTTATTTGTATTTTAAAGAACTTCTAACAAGATTTGCAAGAG	600
QY	1503	AAAGAAATTTGGTGGCCCCCACCCTTCTTAAATTTGGTGACATTAAAA	1548
Db	601	AAAGAAATTTGGTGGCCCCCACCCTTCTTAAATTTGGTGACATTAAAA	646
RESULT 13			
LOCUS	BMT00143	645 bp	mRNA linear EST 28-FEB-2002
DEFINITION	UI-E-DWI-ahc-e-09-0-UI.r1 UI-E-DWI Homo sapiens cDNA clone		
ACCESSION	BMT00143		
VERSION	BMT00143.1	GI:19013401	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 645)		
	Bonaldo,M.F., Lennon,G. and Soares,M.B.		
	Normalization and subtraction: two approaches to facilitate gene		
	discovery		
JOURNAL	Genome Res. 6 (9), 791-806 (1996)		
MEDLINE	9704447		
COMMENT	Contact: Soares, MB		
	Program for Rat Gene Discovery and Mapping		
	University of Iowa		
	451 Eckstein Medical Research Building Iowa City, IA 52242, USA		
	Tel: 319 335 8250		
	Fax: 319 335 9365		
	Email: msocares@blue.weeg.uiowa.edu		
	Tissue Procurement: Dr. Gregg Hageman		
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa		
	Clone Distribution: Researchers may obtain clones from Research		
	Genetics (www.resgen.com).		
	Seq primer: M13 Reverse.		

FEATURES	Source	Location/Qualifiers
		1..645
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="UI-E-DWI-ahc-e-09-0-UI"
		/clone_1lib="UI-E-DWI"
		/tissue-type="lens"
		/dev_stage="adult"
		/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
		/note="Organ: eye; Vector: pT73-pac (Pharmacia) with a modified polylinker; Site_1: Ecor I; Site_2: Not I; UI-E-DWI is a normalized cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT73-pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (drr)18 tail. The sequence tag for this library is CGATTACGA. This library was created for the program, gene discovery in the Visual System, supported by National Eye Institute (NEI)."
BASE COUNT	172 a 190 c 171 g 111 t	1 others
ORIGIN		
Query Match	9.8%	Score 642.4; DB 14; Length 645;
Best Local Similarity	99.7%	Pred. No. 1.7e-119;
Matches 643; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
QY 4007	GAGCCTTGAGAACAGCCACAGTCCCGACAGCTGGGGCTCTTCCACAGGTTACTGCGGAGAGA	4066
Db 1	GAGCCTTGAGAACAGCCACAGTCCCGACAGCTGGGGCTCTTCCACAGGTTACTGCGGAGAGA	60
QY 4067	AGACTCAAAAGGTGACACAGGACACATTAAGCGGAGGGGTGAAAGATGTTTCATTGA	4126
Db 61	AGACTCAAAAGGTGACACAGGACACATTAAGCGGAGGGGTGAAAGATGTTTCATTGA	120
QY 4127	AGCGCAAGCAGTAGACCTTAAGGCTGTGACTACGAAACAAACCAAGCCGTGCCCATGCC	4186
Db 121	AGCGCAAGCAGTAGACCTTAAGGCTGTGACTACGAAACCAAGCCGTGCCCATGCC	180
QY 4187	TGCCACATAGCTGTGGCATCAAGTACTACAAAGGGGCTATTGACGAGAAAGAGGGCAG	4246
Db 181	TGCCACATAGCTGTGGCATCAAGTACTACAAAGGGGCTATTGACGAGAAAGAGGGCAG	240
QY 4247	GTAATGAGAGGCCCCCGCCACCCCTCCGGGCTACTATTGGAATTTCCATTACTAGATTCC	4306
Db 241	GTAATGAGAGGCCCCCGCCACCCCTCCGGGCTACTATTGGAATTTCCATTACTAGATTCC	300
QY 4307	AGAAAGGACACTCCATCCAGCCAGGACGAGGACGCGCGGAGCTACAAAGTGGGCTTCAGAGATC	4366
Db 301	AGAAAGGACACTCCATCCAGCCAGGACGAGGACGCGCGGAGCTACAAAGTGGGCTTCAGAGATC	360
QY 4367	GCGGATGGTCCGACGATCTCTCCGACACAGCTGTGGGCTTCATCCGTACAGACGCCCATGG	4426
Db 361	GCGGATGGTCCGACGATCTCTCCGACACAGCTGTGGGCTTCATCCGTACAGACGCCCATGG	420
QY 4427	GCATCCCAACAGCAGAGCCGCTGAACAAACCTCAGTGGCTTAACAGCAAGCAGATCTGA	4486
Db 421	GCATCCCAACAGCAGAGCCGCTGAACAAACCTCAGTGGCTTAACAGCAAGCAGATCTGA	480
QY 4487	CCCGGCGCTCGCCCCCTTATCACTGCCAAGGGTTTCCACCGAGAGAGATGAAGATGAACA	4546
Db 481	CCCGGCGCTCGCCCCCTTATCACTGCCAAGGGTTTCCACCGAGAGAGATGAAGATGAACA	540
QY 4547	AGTTCTCTCTGTTGAGGACACAGATTTTCTGGAGACAGAGCGAGCCACCTGAAGAGAGA	4606
Db 541	AGTTCTCTCTGTTGAGGACACAGATTTTCTGGAGACAGAGCGAGCCACCTGAAGAGAGA	600
QY 4607	GCACAAGAGAGCTCTGAGCATTTGGAGCTTTGGAACTACATATTC	4651

Db 601 GCACANGAGAGCTCTGACCATTTGGAGCCTTGGAACTGCATTC 645

RESULT 14
AM467230/c 643 bp mRNA linear EST 24-FEB-2000
LOCUS he07906.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2918362 3'
DEFINITION similar to TR:043386 043386 KIA0011B ;, mRNA sequence.
ACCESSION AM467230
VERSION AM467230.1 GI:7037336
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 643)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov
Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange,
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrp/image/image.html
Seq primer: -40UP from Glibco
High quality sequence stop: 391.
Location/Qualifiers
1. 643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2918362"
/clone_lib="NCI_CGAP_CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: pCMV-SPORT6; Site:1:
SalI; Site:2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Library constructed by Life Technologies."
BASE COUNT 152 a 172 c 129 g 190 t
ORIGIN
Query Match 9.5%; Score 627; DB 10; Length 643;
Best Local Similarity 98.4%; Pred. No. 2.2e-116;
Matches 633; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 147 CATCTTACTGACAGTCTCCACCCACAGAGTACCCAGCTTTCTTCTAGCCATTGAGATGT 206
DB 643 CATCTTACTGACAGTCTCCACCCACAGAGTACCCAGCTTTCTTCTAGCCATTGAGATGT 584
QY 207 AGTATCACTAGTATGTTCTGGAGAGAGTCTTCTGATATCTACAGGCGCACAGAAAG 266
DB 583 AGTATCACTAGTATGTTCTGGAGAGAGTCTTCTGATATCTACAGGCGCACAGAAAG 524
QY 267 GAGCGTGTGATATGAGACCTGAGTGGGTGGCCAGAAACAGCAGTGGATTCGGAAGACGAC 326
DB 523 GAGCGTGTGATATGAGACCTGAGTGGGTGGCCAGAAACAGCAGTGGATTCGGAAGACGAC 464
QY 327 GACGATGAAGAAGACATTGAGAGAGCATCAGATCCTCTGATGAGCAGGAGCATTTGTGAGA 386
DB 463 GACGATGAAGAAGACATTGAGAGAGCATCAGATCCTCTGATGAGCAGGAGCATTTGTGAGA 404
QY 387 GACTGCTTAGAAGACAGCAGTGGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 446
DB 403 GACTGCTTAGAAGACAGCAGTGGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 344
QY 447 TTTATGACACAGTGTGCTGCTTTTGGCCAAATATGACAAATGTGATGAGGCGAAGAACTCTGT 506

Db 343 TTTATGACACAGTGTGCTGCTTTTGGCCAAATATGATGATGATGATGATGATGATGATGAT 284
QY 507 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 566
DB 283 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 224
QY 567 GAGCTGACCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 626
DB 223 GAGCTGACCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 164
QY 627 AAAGCAGAAATACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 686
DB 163 AAAGCAGAAATACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 104
QY 687 TACATGAAGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 746
DB 103 TACATGAAGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44
QY 747 CAAGATTACTGCGTATCTCTCATCATCACTAGTAAAGAACATGTC 789
DB 43 CAAGATTACTGCGTATCTCTCATCATCACTAGTAAAGAACATGTC 1

RESULT 15
AL596512 625 bp mRNA linear EST 14-AUG-2001
LOCUS DKF2P451B190.p1 451 (synonym: hicc1) spinal cord Homo sapiens cDNA
DEFINITION Clone DKF2P451B190 5', mRNA sequence.
ACCESSION AL596512
VERSION AL596512.1 GI:15154208
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 625)
Ostenwald, B., Obermaier, B., Mewes, W., Mewes, H.W., Well, B. and
Wiemann, S.
EST (Ostenwald, B., Obermaier, B., Mewes, H.W., Well, B. and Wiemann
S.)
TITLE Unpublished (2001)
JOURNAL Contact: Ostenwald B
COMMENT MIPs
Am Klopferplatz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de.
sequenced by Medigenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. No sl sequence
available.
This clone (DKF2P451B190) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="DKF2P451B190"
/clone_lib="451 (synonym: hicc1) spinal cord"
/tissue_type="human spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site:1: NotI; Site:2: SalI"
BASE COUNT 212 a 97 c 156 g 159 t 1 others
ORIGIN
Query Match 9.5%; Score 624; DB 9; Length 625;
Best Local Similarity 99.8%; Pred. No. 8.7e-116;
Matches 624; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 676 TGGACAAGAAATACATGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 735
|||||

```
Db      1  TGGACAAAGATACATGAAGAAGTGTAGAGAACAAAGGTGATGACTGCCAGTTTGTCT 60
QY      736  GCATAGCCAGCAGAGATTACTGCCGATTTCTCAATCAGTAGAGAAAGACATGCAAAAAG 795
      |||||||
Db      61  GCATAGCCAGCAGAGATTACTGCCGATTTCTCAATCAGTAGAGAAAGACATGCAAAAAG 120
QY      796  TTGAAGAGAGAGAGAGATTGTATGTGTAAAGAACCCGAGAACTTGATCGAACTGAA 855
      |||||||
Db      121  TTGAAGAGAGAGAGATTGTATGTGTAAAGAACCCGAGAACTTGATCGAACTGAA 180
QY      856  CAAGAAAGGAGACATGTTCATCAAGGTAAGGTAAGGTTAAACATGCAATTTGTGTG 915
      |||||||
Db      181  CAAGAAAGGAGACATGTTCATCAAGGTAAGGTAAGGTTAAACATGCAATTTGTGTG 240
QY      916  AAGAGCATTCAGTAGTAGATCCACATTCATAGAAAGACTTTCTGTGACCTATAGGACTT 975
      |||||||
Db      241  AAGAGCATTCAGTAGTAGATCCACATTCATAGAAAGACTTTCTGTGACCTATAGGACTT 300
QY      976  TTCTTTCTAGCCCAATGAAAGTGGGCAAAAAGTTATTGAGTGTGTTAAATGACCCGAGCC 1035
      |||||||
Db      301  TTCTTTCTAGCCCAATGAAAGTGGGCAAAAAGTTATTGAGTGTGTTAAATGACCCGAGCC 360
QY      1036  TCAGGATTAAGGTTACACGGGTAGTATTAATGTGGGTAATAATCACTTCATGACTTGG 1095
      |||||||
Db      361  TCAGGATTAAGGTTACACGGGTAGTATTAATGTGGGTAATAATCACTTCATGACTTGG 420
QY      1096  AAGGAGATCTCGAATGACTCGATTTTGTAGAGAAATTTGAAACAACTCTGGAAGAGAGA 1155
      |||||||
Db      421  AAGGAGATCTCGAATGACTCGATTTTGTAGAGAAATTTGAAACAACTCTGGAAGAGAGA 480
QY      1156  AATGGGTGGACACCTAAGGCTGTTGAATATGCGCTGTGCTGCTAAAGCAAAAAGAGAT 1215
      |||||||
Db      481  AATGGGTGGACACCTAAGGCTGTTGAATATGCGCTGTGCTGCTAAAGCAAAAAGAGAT 540
QY      1216  TGATGACGTTTAACAAACCATCCCGAAGCTCTTGCCTTTATCTTACTTGAAGGCT 1275
      |||||||
Db      541  TGATGACGTTTAACAAACCATCCCGAAGCTCTTGCCTTTATCTTACTTGAAGGCT 600
QY      1276  CTGAGAAGGAGATTGGAATCTTTGT 1300
      |||||||
Db      601  CTGAGAAGGAGATTGGAATCTTTGT 625
```

Search completed: February 18, 2003, 17:34:17
Job time : 7868 secs

BLANK PAGE

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 15:52:41 ; Search time 330 Seconds
(without alignments)
10137.084 Million cell updates/sec

Title: US-09-911-826a-1

Perfect score: 6568

Sequence: 1 cctgcacatcgtagagattg.....aaataataataataacat 6568

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications -NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCT05_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6568	100.0	6568	10	US-09-880-107-1535
2	6568	100.0	6568	10	US-09-911-826a-1
3	1587	24.2	6383	10	US-09-728-952-68
4	399	6.1	399	9	US-10-046-935-1879
5	399	6.1	399	9	US-09-878-178-1879
6	394.8	6.0	799	10	US-09-911-826a-3
7	388	5.9	400	9	US-10-046-935-1196
8	388	5.9	400	9	US-09-878-178-1196
9	227.8	3.5	263	10	US-09-733-607-15
10	201.6	3.1	360	10	US-09-864-761-27564
11	166.6	2.5	465	10	US-09-864-761-10928
12	104.6	1.6	471	10	US-09-864-761-5303
13	104.6	1.6	5900	10	US-09-954-456-1970
14	95.8	1.5	2536	9	US-09-822-846-129
15	87.8	1.3	182	10	US-09-864-761-22078
16	72.8	1.1	181	10	US-09-864-761-31466
17	72.8	1.1	463	10	US-09-864-761-14576
18	60.8	0.9	4770	10	US-09-801-368-63
19	56.2	0.9	446	10	US-09-864-761-11216

C 20	52.4	0.8	383	10	US-09-864-761-28016	Sequence 28016, A
C 21	50	0.8	6200	9	US-09-993-241-1	Sequence 1, Appli
C 22	50	0.8	6200	10	US-09-993-241-1	Sequence 1, Appli
C 23	49.4	0.8	311	10	US-09-983-965-1973	Sequence 1973, Ap
C 24	49.2	0.7	2797	9	US-10-078-770-133	Sequence 133, App
C 25	49	0.7	1178	9	US-09-764-868-410	Sequence 410, App
C 26	49	0.7	1926	10	US-09-764-853-238	Sequence 238, App
C 27	49	0.7	2086	10	US-09-728-952-74	Sequence 74, Appl
C 28	49	0.7	6336	10	US-09-964-824a-114	Sequence 114, App
C 29	49	0.7	6336	10	US-09-880-107-1537	Sequence 1537, Ap
C 30	48.2	0.7	317	10	US-09-960-352-14196	Sequence 14196, A
C 31	48.2	0.7	1586	10	US-09-745-763-86	Sequence 86, Appl
C 32	48	0.7	3275	9	US-09-854-133-151	Sequence 151, App
C 33	48	0.7	3375	10	US-09-738-973-151	Sequence 151, App
C 34	47.6	0.7	3375	10	US-09-764-846-100	Sequence 100, App
C 35	47.2	0.7	597	10	US-09-925-300-87	Sequence 87, Appl
C 36	47	0.7	545	10	US-09-764-873-12	Sequence 12, Appl
C 37	47	0.7	6200	9	US-10-206-443-1	Sequence 1, Appli
C 38	46.8	0.7	401	10	US-09-960-352-14060	Sequence 14060, A
C 39	46.8	0.7	416	10	US-09-960-352-4584	Sequence 4584, A
C 40	46.6	0.7	332	10	US-09-960-352-13112	Sequence 13112, A
C 41	46.6	0.7	1419	10	US-09-940-836a-3	Sequence 3, Appli
C 42	46.6	0.7	2807	10	US-09-940-836a-1	Sequence 1, Appli
C 43	46.4	0.7	203	10	US-09-960-352-3742	Sequence 3742, Ap
C 44	46.4	0.7	392	10	US-09-960-352-2634	Sequence 2634, Ap
C 45	46.2	0.7	219	10	US-09-864-761-27841	Sequence 27841, A

ALIGNMENTS

RESULT 1
US-09-880-107-1535
; Sequence 1535, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherff, Uwe.
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1535
; LENGTH: 6568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AB002311
US-09-880-107-1535

Query Match 100.0%; Score 6568; DB 10; Length 6568;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTGCCATCGTAGAGATTGGTACATGATGTGTAATTCAGTTCAGCATATGTTCTTCA 60
DB 1 CTTGCCATCGTAGAGATTGGTACATGATGTGTAATTCAGTTCAGCATATGTTCTTCA 60
QY 61 TTATGAACACATGACATCCAGCTAACATGAGATTATGGCCAGCAGAGAAACT 120
DB 61 TTATGAACACATGACATCCAGCTAACATGAGATTATGGCCAGCAGAGAAACT 120
QY 121 CACTTCCTGAGATTTCACAAACTGACATCTTACTGACACTTCCACCCAGGTGACCC 180
DB 121 CACTTCCTGAGATTTCACAAACTGACATCTTACTGACACTTCCACCCAGGTGACCC 180

181 ACGTTCTTAGCCATTTCAGATGTAGTATCACTAGTATTCCTGGGACGACAGCTCTT 240
181 ACGTTCTTAGCCATTTCAGATGTAGTATCACTAGTATTCCTGGGACGACAGCTCTT 240
241 CTGATATCTACAGCCACAGAAAGCAGCTGGTATATGACACCTGAGGGCTTCCAG 300
241 CTGATATCTACAGCCACAGAAAGCAGCTGGTATATGACACCTGAGGGCTTCCAG 300
301 AAACAGCAGTGGATTCGGAAGCAGCAGATGAAAGACATTTGAGAGCATTCAGATC 360
301 AAACAGCAGTGGATTCGGAAGCAGCAGATGAAAGACATTTGAGAGCATTCAGATC 360
361 CTCTGATGAGCAGGACATTTGTGAGAGACCTGCTAGAGAGAGACCCCAATTGACCAGCAG 420
361 CTCTGATGAGCAGGACATTTGTGAGAGACCTGCTAGAGAGAGACCCCAATTGACCAGCAG 420
421 ATGATGACATTTGAACACTCTTGGATTTATGCACAGTTGCCCTGCTTTGCCAATTTGA 480
421 ATGATGACATTTGAACACTCTTGGATTTATGCACAGTTGCCCTGCTTTGCCAATTTGA 480
481 CAATGTCAGTGAAGCAGAACTCTGCTGTGATGCTTTGGCAGTGGTGAAGAGCAG 540
481 CAATGTCAGTGAAGCAGAACTCTGCTGTGATGCTTTGGCAGTGGTGAAGAGCAG 540
541 GGACCTATGTTAATGATGTTGTAAGAGCTGCTGCTGCTAGTGAATTTCTCAATGAT 600
541 GGACCTATGTTAATGATGTTGTAAGAGCTGCTGCTGCTAGTGAATTTCTCAATGAT 600
601 CTGTGAGAGTGAATTTATCCAGATGGAAGAGCAAAATACTGTGCATGGGAAATGTTTG 660
601 CTGTGAGAGTGAATTTATCCAGATGGAAGAGCAAAATACTGTGCATGGGAAATGTTTG 660
661 GTGTCTCTCTACCATGAGCAAAAGATATCATGAAAGAGTGAATGAGCAAAAGTGAATG 720
661 GTGTCTCTCTACCATGAGCAAAAGATATCATGAAAGAGTGAATGAGCAAAAGTGAATG 720
721 ACTGCCAGTTTGTCTGCATAGCCAGCAAGATTTACTGCCGTATTTCTCATCAAGTAGAA 780
721 ACTGCCAGTTTGTCTGCATAGCCAGCAAGATTTACTGCCGTATTTCTCATCAAGTAGAA 780
781 AGAATCATGCAAAAGTTGAAAGAGAGAGATTTGTTAGTGAAGAGCAACCCGAGAAC 840
781 AGAATCATGCAAAAGTTGAAAGAGAGAGATTTGTTAGTGAAGAGCAACCCGAGAAC 840
841 TTGATCGAATCGAAGAAAGGAGACATTTGTCATCAAGGTTACTCAGAAAGGTTAA 900
841 TTGATCGAATCGAAGAAAGGAGACATTTGTCATCAAGGTTACTCAGAAAGGTTAA 900
901 CAATGCAATTTGGTGAAGAGCATTCAGTATGATCCAAATTCATATGAGAGACTTTCTGT 960
901 CAATGCAATTTGGTGAAGAGCATTCAGTATGATCCAAATTCATATGAGAGACTTTCTGT 960
961 TGACCTATAGAGCTTTCTTTCTAGCCCAATGAGAGTGGGCAAAATTTATGAGAGGCT 1020
961 TGACCTATAGAGCTTTCTTTCTAGCCCAATGAGAGTGGGCAAAATTTATGAGAGGCT 1020
1021 TTAATGACCCGAGCTCAGGATTAAGTTTACAGGGTAGTTATTTGTTGGTAAATATC 1080
1021 TTAATGACCCGAGCTCAGGATTAAGTTTACAGGGTAGTTATTTGTTGGTAAATATC 1080
1081 ACTTCAATGACTTTGAGAGAGATCTCTGCAATGATCGATTTTGAAGAAATTTGAAACA 1140
1081 ACTTCAATGACTTTGAGAGAGATCTCTGCAATGATCGATTTTGAAGAAATTTGAAACA 1140
1141 ATCTGGAAGAGAGAAATGGTGCACACTAGAGCTTTGAATATGCGGCTGCTGCTA 1200
1141 ATCTGGAAGAGAGAAATGGTGCACACTAGAGCTTTGAATATGCGGCTGCTGCTA 1200
1201 AAGCAAAAGAGATGATGATTAACAAACCATCCCGAGAGCTCTTGGCTTTTA 1260
1201 AAGCAAAAGAGATGATGATTAACAAACCATCCCGAGAGCTCTTGGCTTTTA 1260

1261 TCTTACTTGGAGGCTCTGAGAGGGATTTGGAATCTTTGTGACATGATTCAGSTA 1320
1261 TCTTACTTGGAGGCTCTGAGAGGGATTTGGAATCTTTGTGACATGATTCAGSTA 1320
1321 GCAAGCAACTGAAAGCAGGCTTGAACGGGGGATGATATTTAGAAAGTAATGCGCAAA 1380
1321 GCAAGCAACTGAAAGCAGGCTTGAACGGGGGATGATATTTAGAAAGTAATGCGCAAA 1380
1381 ACTTTGAAACATTCAGCTGTCAAAAGCTATGGAATTTCTTGAATTAACACATTTAT 1440
1381 ACTTTGAAACATTCAGCTGTCAAAAGCTATGGAATTTCTTGAATTAACACATTTAT 1440
1441 CTATCAGCTGAAACCAATTTATTTGATTTAAGAACTTTAAGAAATTTGTCAGAG 1500
1441 CTATCAGCTGAAACCAATTTATTTGATTTAAGAACTTTAAGAAATTTGTCAGAG 1500
1501 AGAAAGAAATGGTGGCCCCCCTCTTAAATTTGTTGATTAATAAGGCGAGCTGCT 1560
1501 AGAAAGAAATGGTGGCCCCCCTCTTAAATTTGTTGATTAATAAGGCGAGCTGCT 1560
1561 ACTCCATTCAGATCTGCTGATGATGAGAGCAGTGTAGAGACTTGAAGAAAGTGAAC 1620
1561 ACTCCATTCAGATCTGCTGATGATGAGAGCAGTGTAGAGACTTGAAGAAAGTGAAC 1620
1621 AAAAAAGTAAAGCCAACTGTGGAGAGAGCAACAGCTGCAAAAGATCTGAGACAA 1680
1621 AAAAAAGTAAAGCCAACTGTGGAGAGAGCAACAGCTGCAAAAGATCTGAGACAA 1680
1681 CTGCGATCAGATCTTGGCCAGAGAAACATGATGATTTGGATGGATGCTCAGTCAAG 1740
1681 CTGCGATCAGATCTTGGCCAGAGAAACATGATGATTTGGATGGATGCTCAGTCAAG 1740
1741 ATGACAGCATAGTAGATTAAGGAGAGAGCAAAAGCAGATCCCAATGCTGCTGCTG 1800
1741 ATGACAGCATAGTAGATTAAGGAGAGAGCAAAAGCAGATCCCAATGCTGCTGCTG 1800
1801 GAACCTTATCATCATTAATCCGATTTATTTGACATGATGATGATTTTGAACCTTCA 1860
1801 GAACCTTATCATCATTAATCCGATTTATTTGACATGATGATGATTTTGAACCTTCA 1860
1861 GTGCTACTCTGCTGCTGCTGATCAAGTCAAGTCAAGGCTTTTAAAGCTATCAAGAACCC 1920
1861 GTGCTACTCTGCTGCTGCTGATCAAGTCAAGTCAAGGCTTTTAAAGCTATCAAGAACCC 1920
1921 GCTTATCATGATCAGTAAGAGACACTACAGCAAGAGAGATGCTATGAGCTATCAGGG 1980
1921 GCTTATCATGATCAGTAAGAGACACTACAGCAAGAGAGATGCTATGAGCTATCAGGG 1980
1981 AGTTGCTGTTACTGCCACCCCGATCAATATTTCAATGTGAGCTGTGTCACACCTG 2040
1981 AGTTGCTGTTACTGCCACCCCGATCAATATTTCAATGTGAGCTGTGTCACACCTG 2040
2041 AGGAGATATCAAAAGAAAGAGCTTCCAGATGAGTCAAGCTTTCCAAACTTGCAGAGAAATAC 2100
2041 AGGAGATATCAAAAGAAAGAGCTTCCAGATGAGTCAAGCTTTCCAAACTTGCAGAGAAATAC 2100
2101 AACGAGGAGAGGATATCTGAAAGAAACATGAAAGAGAACTTTGTTGATGATG 2160
2101 AACGAGGAGAGGATATCTGAAAGAAACATGAAAGAGAACTTTGTTGATGATG 2160
2161 AAGATGCTCAGAGCTTTGTTGAGAGAGTCAAAATTTCCCTCTTCAAGCTCAGAGCTG 2220
2161 AAGATGCTCAGAGCTTTGTTGAGAGAGTCAAAATTTCCCTCTTCAAGCTCAGAGCTG 2220
2221 AAGTTGCAACACAGCTCTCTATGCAAAATTTTGAACCTTTGCAACATTTGAACCTG 2280
2221 AAGTTGCAACACAGCTCTCTATGCAAAATTTTGAACCTTTGCAACATTTGAACCTG 2280
2281 AATATATGATGATTTATTTAACTCAGATCAAAACAGCTGTGCGCAACCTGGAAGAT 2340
2281 AATATATGATGATTTATTTAACTCAGATCAAAACAGCTGTGCGCAACCTGGAAGAT 2340
2341 TTGAGAGAGTCAATTAACGAGAAACATTTTGGGTAGATCTGAAATTTCTCAGAGAAACA 2400

|||||
Db 2341 TTGAAAGATCATTAACCGAGAAACATTTGGGTAGCATCTGAATTTTCAGAGAAACAA 2400
QY 2401 ACCAGCTGAAGAGATGAGAGATCATTAAGCATTTTCATCAAGATAGACATGCACTGAGGG 2460
Db 2401 ACCAGCTGAAGAGATGAGAGATCATTAAGCATTTTCATCAAGATAGACATGCACTGAGGG 2460
QY 2461 AATGCAAGAAATTTAACTCAATGTTTGCAATCATCAAGTGGCCCTAAACCTGGGACCAAGTGG 2520
Db 2461 AATGCAAGAAATTTAACTCAATGTTTGCAATCATCAAGTGGCCCTAAACCTGGGACCAAGTGG 2520
QY 2521 CAAGACTGGCAACGACCTGGGAGAAACCTTCCCAATAAATACGAAAACTATTTCAAGATC 2580
Db 2521 CAAGACTGGCAACGACCTGGGAGAAACCTTCCCAATAAATACGAAAACTATTTCAAGATC 2580
QY 2581 TCCAAACACCTGTTGATCTCTCCAGAAACATGCGCAAAATATGCTATGTTCTCAATAGTC 2640
Db 2581 TCCAAACACCTGTTGATCTCTCCAGAAACATGCGCAAAATATGCTATGTTCTCAATAGTC 2640
QY 2641 AAAATCTACACCTCCCATATCCCTCTATTTCCAGTTATCAAAAAAGGATCTCACCTTCC 2700
Db 2641 AAAATCTACACCTCCCATATCCCTCTATTTCCAGTTATCAAAAAAGGATCTCACCTTCC 2700
QY 2701 TTCACGAAGAAATGACTCAAAAAGTAGACGGCTGCTCAATTTTGAAGCTAAGAGATGA 2760
Db 2701 TTCACGAAGAAATGACTCAAAAAGTAGACGGCTGCTCAATTTTGAAGCTAAGAGATGA 2760
QY 2761 TTGCAAAAGAAATTTGCTCAGCTGGCCGAATGCTTCAGTGAACATGACCCCTGCTCA 2820
Db 2761 TTGCAAAAGAAATTTGCTCAGCTGGCCGAATGCTTCAGTGAACATGACCCCTGCTCA 2820
QY 2821 TGTTCAGAGACTCGGAAGAAATGGGAGATTGGGGTCTCTCAGCAGAGGTAGTACAA 2880
Db 2821 TGTTCAGAGACTCGGAAGAAATGGGAGATTGGGGTCTCTCAGCAGAGGTAGTACAA 2880
QY 2881 ATGCAACAGTGTAGATGTTGCTCAGACAGGTGTCATAAAAAAGCGGTACGTAGTT 2940
Db 2881 ATGCAACAGTGTAGATGTTGCTCAGACAGGTGTCATAAAAAAGCGGTACGTAGTT 2940
QY 2941 CCTTTTCATATGCCAAAAAAGCTTTATGAGATATGCCAAATGGCTCGAAAAAGTACAGCAT 3000
Db 2941 CCTTTTCATATGCCAAAAAAGCTTTATGAGATATGCCAAATGGCTCGAAAAAGTACAGCAT 3000
QY 3001 ACCTTTCAATTTGAGACTGAGAAATGAGACAGAGAGTCTTCAGACATTTTCTCTGAGT 3060
Db 3001 ACCTTTCAATTTGAGACTGAGAAATGAGACAGAGAGTCTTCAGACATTTTCTCTGAGT 3060
QY 3061 GTGAGCCAGCAACCAACACATTTGCTTAAGATCTGTGTGACAAAAAGCCTGTCAAAATCCG 3120
Db 3061 GTGAGCCAGCAACCAACACATTTGCTTAAGATCTGTGTGACAAAAAGCCTGTCAAAATCCG 3120
QY 3121 AGACCTTCCAGTAGTCCCAAGGGGAGGTCACAAACAGAAAGTCCAGTCCCTGCCACAGC 3180
Db 3121 AGACCTTCCAGTAGTCCCAAGGGGAGGTCACAAACAGAAAGTCCAGTCCCTGCCACAGC 3180
QY 3181 CCCAGAGCAGCCACCACACGACATAAATCAACAGAGGATACAGGTTCCGCCGCTGT 3240
Db 3181 CCCAGAGCAGCCACCACACGACATAAATCAACAGAGGATACAGGTTCCGCCGCTGT 3240
QY 3241 CCTTTATCTTCAGGAGAAAGTGCCGTAAAGATCTCCGACCTTTTGGCATAACT 3300
Db 3241 CCTTTATCTTCAGGAGAAAGTGCCGTAAAGATCTCCGACCTTTTGGCATAACT 3300
QY 3301 CTTCACACAGTTTAAAAAAATCTTTCTTGTCTGAAGAGAGATTTTGAACGTCACA 3360
Db 3301 CTTCACACAGTTTAAAAAAATCTTTCTTGTCTGAAGAGAGATTTTGAACGTCACA 3360
QY 3361 AGAAACAGGCTGAAGATACATATCAAAATGCAATCTTCGACGCTTTCTTCCTCCTACTT 3420
Db 3361 AGAAACAGGCTGAAGATACATATCAAAATGCAATCTTCGACGCTTTCTTCCTCCTACTT 3420
QY 3421 CTCACAGAGTTTCCAGAGAAAGGCTATCTTTGGCTCCAGTGTACTGTGATTAAT 3480
|||||

Db 3421 CTCACAGAGTTTCCAGAGAAAGCCTATCTTTGGCTCCACAGTGTACTGTGATTAAT 3480
QY 3481 TTTCACATTTCTGTCTACAGTGAATTTTCTTCACAGATCCAGTATTTGTAGCAATTCGCTT 3540
Db 3481 TTTCACATTTCTGTCTACAGTGAATTTTCTTCACAGATCCAGTATTTGTAGCAATTCGCTT 3540
QY 3541 TTGACATCAGTGCAGTCTCACTGACAGATGAGAGGGCCGAGAGGCAATTTGTCAGATG 3600
Db 3541 TTGACATCAGTGCAGTCTCACTGACAGATGAGAGGGCCGAGAGGCAATTTGTCAGATG 3600
QY 3601 TGGAAACAAACCTTAGGAGATGGGAGAGATGAGAGGGGACCATGATTAACCTGATCAGT 3660
Db 3601 TGGAAACAAACCTTAGGAGATGGGAGAGATGAGAGGGGACCATGATTAACCTGATCAGT 3660
QY 3661 ATAGCTTGGGGTCTTATGACCAATGTCCGAGGGCCGAGGCTTATATGCTACAGCTACAG 3720
Db 3661 ATAGCTTGGGGTCTTATGACCAATGTCCGAGGGCCGAGGCTTATATGCTACAGCTACAG 3720
QY 3721 TAAATCTTCTCCACAGCAGACAGAGACTTTCCAGAGATCAGGGGATCGCGCTCACCTG 3780
Db 3721 TAAATCTTCTCCACAGCAGACAGAGAACTTTCCAGAGATCAGGGGATCGCGCTCACCTG 3780
QY 3781 ATGCTGCTGACAGTGGCCGTGGAGCTGACGTCATGCTCAAGTGGCTCCCATGATTAATA 3840
Db 3781 ATGCTGCTGACAGTGGCCGTGGAGCTGACGTCATGCTCAAGTGGCTCCCATGATTAATA 3840
QY 3841 TACAGAGATCCAGCAGACAGAGAAAGCTGGGAGACTCTTCATTTGGGCAATCTGCTTG 3900
Db 3841 TACAGAGATCCAGCAGACAGAGAAAGCTGGGAGACTCTTCATTTGGGCAATCTGCTTG 3900
QY 3901 ATTATTCAGGGAGTCTGCGAGTTTATGGGCATCAAGCAGCCATATGAGCAAAATATGT 3960
Db 3901 ATTATTCAGGGAGTCTGCGAGTTTATGGGCATCAAGCAGCCATATGAGCAAAATATGT 3960
QY 3961 TTTCGTATATACACAAAGTATACAGGCAAAATCAAAAGTACAGAGAGCCTTGAACAG 4020
Db 3961 TTTCGTATATACACAAAGTATACAGGCAAAATCAAAAGTACAGAGAGCCTTGAACAG 4020
QY 4021 CCCAGTCCGAGCAAGCTGGGCTCTTCCACAGGTTACTGGGAGAAAGACTCGAAAGGTG 4080
Db 4021 CCCAGTCCGAGCAAGCTGGGCTCTTCCACAGGTTACTGGGAGAAAGACTCGAAAGGTG 4080
QY 4081 ACACAGCACAATTAAGCGGAGGGTGAAGAGATGTTTCCATTGAAACCGGAAGCAGTA 4140
Db 4081 ACACAGCACAATTAAGCGGAGGGTGAAGAGATGTTTCCATTGAAACCGGAAGCAGTA 4140
QY 4141 GCCTAAAGTCTGTGACTACGGAAGAAACCAAGCTGTGCCATGCCCTGCACATAGCTG 4200
Db 4141 GCCTAAAGTCTGTGACTACGGAAGAAACCAAGCTGTGCCATGCCCTGCACATAGCTG 4200
QY 4201 TGGCATCAAGTACTACAAAGGGGCTATTGCAAGAAAGGAGGCTATGAGAGACCCG 4260
Db 4201 TGGCATCAAGTACTACAAAGGGGCTATTGCAAGAAAGGAGGCTATGAGAGACCCG 4260
QY 4261 CGCCCAACCCCTCCGGCTACATTTGGAATTTCCATTAAGTGAAGAGGACACTCC 4320
Db 4261 CGCCCAACCCCTCCGGCTACATTTGGAATTTCCATTAAGTGAAGAGGACACTCC 4320
QY 4321 ATCCAGCCAGAAACCGCCGACTACAAAGTGGCCCTTCAGAGATCGCGATGCTGCAC 4380
Db 4321 ATCCAGCCAGAAACCGCCGACTACAAAGTGGCCCTTCAGAGATCGCGATGCTGCAC 4380
QY 4381 GATTCCTCGACACAGTGGGCTTTCATCCGTACAGACGCCACATGGGCATCCACACAGCA 4440
Db 4381 GATTCCTCGACACAGTGGGCTTTCATCCGTACAGACGCCACATGGGCATCCACACAGCA 4440
QY 4441 GCAGGCTGTGAACAAACCTCAGTGGCATTAACCGAAGCGTGGACCCGCTGGCC 4500
Db 4441 GCAGGCTGTGAACAAACCTCAGTGGCATTAACCGAAGCGTGGACCCGCTGGCC 4500
QY 4501 CTTATAGTCCCAAGGATTTTCCACGAGAGAGATGAAGATGAACAAGTTTGTGCTGTTT 4560
Db 4501 CTTATAGTCCCAAGGATTTTCCACGAGAGAGATGAAGATGAACAAGTTTGTGCTGTTT 4560

QY 4561 GAGGCACAGACTTTCTGGAAGCAGAGCCACCCTGAAGAAGAGACACAAAGACGT 4620
|||||
Db 4561 GAGGCACAGACTTTCTGGAAGCAGAGCCACCCTGAAGAAGAGAGCACAAGACGT 4620
QY 4621 CCTGAGCATTTGGAGCCTTTGGAACCTACATTTCTGAGAGCGGTGAGACAGTTGGCTCTTC 4680
|||||
Db 4621 CCTGAGCATTTGGAGCCTTTGGAACCTACATTTCTGAGAGCGGTGAGACAGTTGGCTCTTC 4680
QY 4681 CCTGCTTTAAAGACAGATGGGGCTCTCTCCCTCTCTCCCTCTCTCCCTGATGTA 4740
|||||
Db 4681 CCTGCTTTAAAGACAGATGGGGCTCTCTCCCTCTCTCCCTCTCTCCCTGATGTA 4740
QY 4741 AATCTGTGAAGAAATGGCTGGACCTTTTCAGACTTTGTGCTTGAATGACAGCTGC 4800
|||||
Db 4741 AATCTGTGAAGAAATGGCTGGACCTTTTCAGACTTTGTGCTTGAATGACAGCTGC 4800
QY 4801 AGCAATCTTGAGCTCCCACTGTTGCTGCTGCGACATCACACAGTATCANTCCAAATTC 4860
|||||
Db 4801 AGCAATCTTGAGCTCCCACTGTTGCTGCTGCGACATCACACAGTATCANTCCAAATTC 4860
QY 4861 CAAGATCTCAACAACAGATGATTCACCTGGCTGACCTTCTCATGGCTGGAAGATT 4920
|||||
Db 4861 CAAGATCTCAACAACAGATGATTCACCTGGCTGACCTTCTCATGGCTGGAAGATT 4920
QY 4921 TTTTAAATCTCTTTAGATTTCATTCACCTAGCTGATCTCATTTGGGATATG 4980
|||||
Db 4921 TTTTAAATCTCTTTAGATTTCATTCACCTAGCTGATCTCATTTGGGATATG 4980
QY 4981 AAAAAAGCAGCATTTGAATCTTGGGGCCCTTAAACCACCAAGAGAGACAAAGAAA 5040
|||||
Db 4981 AAAAAAGCAGCATTTGAATCTTGGGGCCCTTAAACCACCAAGAGAGACAAAGAAA 5040
QY 5041 CAATGAATCTCTTGAATACAGTCTTGCACCTGTTTACAGTCTCTCTTTAAAA 5100
|||||
Db 5041 CAATGAATCTCTTGAATACAGTCTTGCACCTGTTTACAGTCTCTCTTTAAAA 5100
QY 5101 AAAAAAGCTTTAAAGATTTTCTCAGAGAGTAATATATATCCATTATGATTAAG 5160
|||||
Db 5101 AAAAAAGCTTTAAAGATTTTCTCAGAGAGTAATATATATATCCATTATGATTAAG 5160
QY 5161 TTTATTTAAACCTTAAGTAGGGTTGCCAGCGTTCCTGAAAAACCAATATGCCGG 5220
|||||
Db 5161 TTTATTTAAACCTTAAGTAGGGTTGCCAGCGTTCCTGAAAAACCAATATGCCGG 5220
QY 5221 ACAGGGTGTGGCCACACCAAGAGAGGGAAGACTGCTTGTGACCTGGCTCCATG 5280
|||||
Db 5221 ACAGGGTGTGGCCACACCAAGAGAGGGAAGACTGCTTGTGACCTGGCTCCATG 5280
QY 5281 TCCCTGCTGCTACCCCGGAGTGCCTATCTGGAATGTAATGTTACCAATTAA 5340
|||||
Db 5281 TCCCTGCTGCTACCCCGGAGTGCCTATCTGGAATGTAATGTTACCAATTAA 5340
QY 5341 TACCAAGACACCTATCTCTCTCCCAAGTGGATGGGCTTCTGTAACAGTGTG 5400
|||||
Db 5341 TACCAAGACACCTATCTCTCTCCCAAGTGGATGGGCTTCTGTAACAGTGTG 5400
QY 5401 CACATGGCCAGGGGAGGAACTAGACCTTGTCTGCTGAGGCTTATGAGGACAG 5460
|||||
Db 5401 CACATGGCCAGGGGAGGAACTAGACCTTGTCTGCTGAGGCTTATGAGGACAG 5460
QY 5461 ACAGTGTGATGGCGAGTGTCTCTCCATGAGATGATGCAACCCCAATTTTAA 5520
|||||
Db 5461 ACAGTGTGATGGCGAGTGTCTCTCCATGAGATGATGCAACCCCAATTTTAA 5520
QY 5521 GTTATATTTCTTTGATTTGTTAAATTAGAGGTGATTTGTTTTTTTGT 5580
|||||
Db 5521 GTTATATTTCTTTGATTTGTTAAATTAGAGGTGATTTGTTTTTTTGT 5580
QY 5581 TTTTAAAGAGAACTTTATATACGATGATGATGATGAAAGAGAGCTTGGATGTT 5640
|||||
Db 5581 TTTTAAAGAGAACTTTATATACGATGATGATGATGAAAGAGAGCTTGGATGTT 5640

QY 5641 GGAGCTAATGCCAGCTGTTATACGCTCTTTCAAGACAGCTCCCTTATTTGATTTGCC 5700
|||||
Db 5641 GGAGCTAATGCCAGCTGTTATACGCTCTTTCAAGACAGCTCCCTTATTTGATTTGCC 5700
QY 5701 ATTAGGAATTAACAAGCTTTTAAACGTGATTAAGATCAAAAACCTGTTAGCATGCC 5760
|||||
Db 5701 ATTAGGAATTAACAAGCTTTTAAACGTGATTAAGATCAAAAACCTGTTAGCATGCC 5760
QY 5761 AGCTTTGCAAGCGAGTGTAGTACACCAAGACTAACCTCAAGTGGCTTATGAGCGTG 5820
|||||
Db 5761 AGCTTTGCAAGCGAGTGTAGTACACCAAGACTAACCTCAAGTGGCTTATGAGCGTG 5820
QY 5821 CATPATAGAGAAGCCCTAAGTGTAGCAACACTGCTCAGCTGATTTAAACCTTAAT 5880
|||||
Db 5821 CATPATAGAGAAGCCCTAAGTGTAGCAACACTGCTCAGCTGATTTAAACCTTAAT 5880
QY 5881 GACTGAATGACCCCTCAGCTGATTTTGTGTTTGTGCAAGACACTCCGGAAAAGTGA 5940
|||||
Db 5881 GACTGAATGACCCCTCAGCTGATTTTGTGTTTGTGCAAGACACTCCGGAAAAGTGA 5940
QY 5941 AGCTGCCAATCGAGTGTAGTACCAATGTGGAAGAGCTGCTGATTTTTC 6000
|||||
Db 5941 AGCTGCCAATCGAGTGTAGTACCAATGTGGAAGAGCTGCTGATTTTTC 6000
QY 6001 ATTAAATCAGCTGATCATTTGATCAGATGATTAACGTAATAGCTTCAATTTTAA 6060
|||||
Db 6001 ATTAAATCAGCTGATCATTTGATCAGATGATTAACGTAATAGCTTCAATTTTAA 6060
QY 6061 GTGGAATGTAGCTTTTCTCAGTGTATCAACAATGTGAGTGTATTTATTAATCT 6120
|||||
Db 6061 GTGGAATGTAGCTTTTCTCAGTGTATCAACAATGTGAGTGTATTTATTAATCT 6120
QY 6121 CTCTGATCATGCAATTTGCTACTTGTCTTATTTACATTTGCAATTTGTAAT 6180
|||||
Db 6121 CTCTGATCATGCAATTTGCTACTTGTCTTATTTACATTTGCAATTTGTAAT 6180
QY 6181 TTACATGTAATGATTAATTTTCCAGTGTATTAATATAGCTATGACATTTAGT 6240
|||||
Db 6181 TTACATGTAATGATTAATTTTCCAGTGTATTAATATAGCTATGACATTTAGT 6240
QY 6241 ATGAAGAAGCAAAATGTACCTGACACAAAGTGCACAAATGTTTCTAAGCATTAAGT 6300
|||||
Db 6241 ATGAAGAAGCAAAATGTACCTGACACAAAGTGCACAAATGTTTCTAAGCATTAAGT 6300
QY 6301 AATGTGAACATGAGACTGCTAATCTCACTGCTGATGTAAGTGAAGTGA 6360
|||||
Db 6301 AATGTGAACATGAGACTGCTAATCTCACTGCTGATGTAAGTGAAGTGA 6360
QY 6361 CAATTAACCTGATTTTCCATACCTTTGATACCTGATGTAAGTGAAGTGA 6420
|||||
Db 6361 CAATTAACCTGATTTTCCATACCTTTGATACCTGATGTAAGTGAAGTGA 6420
QY 6421 AGACATTTGAGTGTGATTTCCCTTATTTTAAATTAATTTGATTTGAT 6480
|||||
Db 6421 AGACATTTGAGTGTGATTTCCCTTATTTTAAATTAATTTGATTTGAT 6480
QY 6481 TATATTTTGTGAAGATGTGAGAAATGTACTATGTTTATGCTTCAATCAGCTTTGTA 6540
|||||
Db 6481 TATATTTTGTGAAGATGTGAGAAATGTACTATGTTTATGCTTCAATCAGCTTTGTA 6540
QY 6541 CAAGCTGGAATAATTAATTAATTAAT 6568
|||||
Db 6541 CAAGCTGGAATAATTAATTAATTAAT 6568

RESULT 2

US-09-911-826a-1
; Sequence 1, Application US/09911826a

; Patent No. US20020143164A1

; GENERAL INFORMATION:

; APPLICANT: Rolin, Danjela and Pham, Nam

; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and

; TITLE OF INVENTION: Methods of Use


```

; FILE REFERENCE: DDM-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 6568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)..(4562)
US-09-911-826a-1
```

Query Match 100.0%; Score 6568; DB 10; Length 6568;

Best Local Similarity 100.0%; Pred. No. 0; Matches 6568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 CTTGCCATCGTGAGAGATTGGTACATGATGTAAATTCAGTTCACGATATGTTCTTCA 60
    |||||||
Db 1 CTTGCCATCGTGAGAGATTGGTACATGATGTAAATTCAGTTCACGATATGTTCTTCA 60
OY 61 TTATGAACACCTAGCAATCCACAGCTAACCATGGAGTTATGGCCGACGAGAAACACT 120
    |||||||
Db 61 TTATGAACACCTAGCAATCCACAGCTAACCATGGAGTTATGGCCGACGAGAAACACT 120
OY 121 CACTTCCTGAGATTTCACAAAACCTCATCTTACTGACAGTCTCCACCCACAGTGACCC 180
    |||||||
Db 121 CACTTCCTGAGATTTCACAAAACCTCATCTTACTGACAGTCTCCACCCACAGTGACCC 180
OY 181 ACCTTTCTTCTAGCCATTCAGATGTATGATCTAGTATCTGAGATCTGGAGCAGACGCTTT 240
    |||||||
Db 181 ACCTTTCTTCTAGCCATTCAGATGTATGATCTAGTATCTGAGATCTGGAGCAGACGCTTT 240
OY 241 CTGATATCTACCGCCACAGAAAGGAGGCTGCTGATATGACCTGAGTGGGTTCGCA 300
    |||||||
Db 241 CTGATATCTACCGCCACAGAAAGGAGGCTGCTGATATGACCTGAGTGGGTTCGCA 300
OY 301 AAACACAGTGGATTCCGAAGACGACGATGAAGAAGACATTGGAGAGCATCGATC 360
    |||||||
Db 301 AAACACAGTGGATTCCGAAGACGACGATGAAGAAGACATTGGAGAGCATCGATC 360
OY 361 CTCTGATGAGCAGGACATTTGTGAGAGACTGCCCTAGAGAAAGCAATTTGACCGGACAG 420
    |||||||
Db 361 CTCTGATGAGCAGGACATTTGTGAGAGACTGCCCTAGAGAAAGCAATTTGACCGGACAG 420
OY 421 ATGATGACATTTGAACAACCTTGGAAATTTATGACCAAGTTGCTGCTTTTGGCAATATGA 480
    |||||||
Db 421 ATGATGACATTTGAACAACCTTGGAAATTTATGACCAAGTTGCTGCTTTTGGCAATATGA 480
OY 481 CAATGTCACTGAGGAGCAAACTCTGCTGTGATGGTGTTCGCAAGTGGTGAAGAAGCAG 540
    |||||||
Db 481 CAATGTCACTGAGGAGCAAACTCTGCTGTGATGGTGTTCGCAAGTGGTGAAGAAGCAG 540
OY 541 GGAACCATAGTAAATGATGTGAAAGCTGACCTCGGTCAAGTATTTCAATGAT 600
    |||||||
Db 541 GGAACCATAGTAAATGATGTGAAAGCTGACCTCGGTCAAGTATTTCAATGAT 600
OY 601 CTCTGGAAGTGAAGCTTTCAGATGAGAAAGCAAAATACTGTGCAATGGGAAATGATTTTG 660
    |||||||
Db 601 CTCTGGAAGTGAAGCTTTCAGATGAGAAAGCAAAATACTGTGCAATGGGAAATGATTTTG 660
OY 661 GTGTCTCTCTCAATGAGCAAAAGATACATGAAGAGATGATGAGAAACAAAGTGGATG 720
    |||||||
Db 661 GTGTCTCTCTCAATGAGCAAAAGATACATGAAGAGATGATGAGAAACAAAGTGGATG 720
OY 721 ACTGCCAGTTTGTCTGATAGCCCAAGAAATTTACTGCGGTATTTCAATCAAGTAGAAA 780
    |||||||
Db 721 ACTGCCAGTTTGTCTGATAGCCCAAGAAATTTACTGCGGTATTTCAATCAAGTAGAAA 780
```

```

OY 781 AGACATGCAAAAAGTTGAAGAGAGAGAGATTGTTATGTTGGAAGAAGACCCAGAGAC 840
    |||||||
Db 781 AGACATGCAAAAAGTTGAAGAGAGAGAGATTGTTATGTTGGAAGAAGACCCAGAGAC 840
OY 841 TTGATGAACCTGGAACAAAGAGGACACATTTGTCATCAAGGGTACCTGAGAAAGTTAA 900
    |||||||
Db 841 TTGATGAACCTGGAACAAAGAGGACACATTTGTCATCAAGGGTACCTGAGAAAGTTAA 900
OY 901 CAATGCATTTGGTGAAGAGCATTCAGTAGTAGATCAACATTCATAGAGACATTTCTGT 960
    |||||||
Db 901 CAATGCATTTGGTGAAGAGCATTCAGTAGTAGATCAACATTCATAGAGACATTTCTGT 960
OY 961 TGACCTATAGGACTTTCTTTCTAGCCCAATGAGATGGGCAAAAAGTTATGAGATGCT 1020
    |||||||
Db 961 TGACCTATAGGACTTTCTTTCTAGCCCAATGAGATGGGCAAAAAGTTATGAGATGCT 1020
OY 1021 TTATATGACCCGAGCCTCAGGAGTAAAGTTACACGGGTAGTATTTATGTTGTTAAATATC 1080
    |||||||
Db 1021 TTATATGACCCGAGCCTCAGGAGTAAAGTTACACGGGTAGTATTTATGTTGTTAAATATC 1080
OY 1081 ACTTCAATGACTTTGAAGAGATCTGCAATGACATCGATTTTGAAGAATTTGAAGAACA 1140
    |||||||
Db 1081 ACTTCAATGACTTTGAAGAGATCTGCAATGACATCGATTTTGAAGAATTTGAAGAACA 1140
OY 1141 ATCTGAAAGAGAGAAATGCGTGACACCTAAGGCTGTTGAATATGCGGTGCTGCTA 1200
    |||||||
Db 1141 ATCTGAAAGAGAGAAATGCGTGACACCTAAGGCTGTTGAATATGCGGTGCTGCTA 1200
OY 1201 AAGCAAAAAGAGATGATGACGTTACAAACCATCCGAGAGAGCTCCCTTGCCCTTTA 1260
    |||||||
Db 1201 AAGCAAAAAGAGATGATGACGTTACAAACCATCCGAGAGAGCTCCCTTGCCCTTTA 1260
OY 1261 TCTTACTTGGAGGCTCTGAAAGGATTTGGAATCTTTGTTGACAGTGTAGATTACGTA 1320
    |||||||
Db 1261 TCTTACTTGGAGGCTCTGAAAGGATTTGGAATCTTTGTTGACAGTGTAGATTACGTA 1320
OY 1321 GCAAACCACTGAGAGAGGCTTGAACGGGGGATCAGATATTAGAAGTAAATGGCCAAA 1380
    |||||||
Db 1321 GCAAACCACTGAGAGAGGCTTGAACGGGGGATCAGATATTAGAAGTAAATGGCCAAA 1380
OY 1381 ACTTGAACCAATTCAGCTGTCAAAAGCTATGGAATTCCTTAGAATTAACACATTTAT 1440
    |||||||
Db 1381 ACTTGAACCAATTCAGCTGTCAAAAGCTATGGAATTCCTTAGAATTAACACATTTAT 1440
OY 1441 CTATCACTGTGAACCAATTTATTTGATTTAAGAACTTCTAACAGATTGTGCAAG 1500
    |||||||
Db 1441 CTATCACTGTGAACCAATTTATTTGATTTAAGAACTTCTAACAGATTGTGCAAG 1500
OY 1501 AGAAAGAAATGCTGCCCCCACCCTTCTTAATTTGCTGACATTAAGAGCCAGTCGCT 1560
    |||||||
Db 1501 AGAAAGAAATGCTGCCCCCACCCTTCTTAATTTGCTGACATTAAGAGCCAGTCGCT 1560
OY 1561 ACTCCATTCACAGATCTTGTGTGATGTAGAAACAGGTAGAGACTTGAAGAAAGTAAACA 1620
    |||||||
Db 1561 ACTCCATTCACAGATCTTGTGTGATGTAGAAACAGGTAGAGACTTGAAGAAAGTAAACA 1620
OY 1621 AAAAAGTAAAGCAACACTGTGGAGAGAAAGCAAGCTGAAAAGATATCTGCACAAGA 1680
    |||||||
Db 1621 AAAAAGTAAAGCAACACTGTGGAGAGAAAGCAAGCTGAAAAGATATCTGCACAAGA 1680
OY 1681 CTCTGATCACTGATCTTGTGCAACAAACCATACATATATTTGGGATTTGGTCACTCAAG 1740
    |||||||
Db 1681 CTCTGATCACTGATCTTGTGCAACAAACCATACATATATTTGGGATTTGGTCACTCAAG 1740
OY 1741 ATGACAGCATATGAGATTAAAGCAGACAAAGCAATCCAACTGCAATGCTGTAGTG 1800
    |||||||
Db 1741 ATGACAGCATATGAGATTAAAGCAGACAAAGCAATCCAACTGCAATGCTGTAGTG 1800
OY 1801 GAACTTATCATTCAGTAATCTGATTTATTCAGTCACATATCGCATTTTGAAGCTTCA 1860
    |||||||
Db 1801 GAACTTATCATTCAGTAATCTGATTTATTCAGTCACATATCGCATTTTGAAGCTTCA 1860
```

QY 1861 GGTGCTCTCTGCTACTTGGCAGATCAAGTCTAAGGGTTTTTAAAGGCTGATCAGCAAAAGCC 1920
DB 1861 GTCTTACTCTGCTACTGCTCCAGATCAAGTGTCTAAGGGTTTTTAAAGGCTGATCAGCAAAAGCC 1920
QY 1921 GGTACATCATGATCATAGTAAAGACATCAGCAAGCAAGAAAGTGGTCTATTCAGGGCTATCAGGG 1980
DB 1921 GGTACATCATGATCATAGTAAAGACATCAGCAAGCAAGAAAGTGGTCTATTCAGGGCTATCAGGG 1980
QY 1981 AGTTTGGTGTACTGGCCACCCTGGATCAATATTCATATGTAGAGTCTCTGTCCACACTG 2040
DB 1981 AGTTTGGTGTACTGGCCACCCTGGATCAATATTCATATGTAGAGTCTCTGTCCACACTG 2040
QY 2041 AGGAGCTATCAAAACAAGAAAGACTTCCAGATCAGCTTTCCAAACTTTGGTTCAGATG 2100
DB 2041 AGGAGCTATCAAAACAAGAAAGACTTCCAGATCAGCTTTCCAAACTTTGGTTCAGATG 2100
QY 2101 AACTGACTGGAAGGTATTATCTGAAAAACAACATGGAAACAGAAAGCTTTGTTTCAGATG 2160
DB 2101 AACTGACTGGAAGGTATTATCTGAAAAACAACATGGAAACAGAAAGCTTTGTTTCAGATG 2160
QY 2161 AAGATGCTCAGGAGTGTGTGAGAGAGTCAAAATTTCCCTCTCTTCAGCTCAGCAGTGTG 2220
DB 2161 AAGATGCTCAGGAGTGTGTGAGAGAGTCAAAATTTCCCTCTCTTCAGCTCAGCAGTGTG 2220
QY 2221 AAGTTGCAACACAGCTCTCTATGCGGAATTTTGAACCTTTTCCGACACTTGAACCTTCTG 2280
DB 2221 AAGTTGCAACACAGCTCTCTATGCGGAATTTTGAACCTTTTCCGACACTTGAACCTTCTG 2280
QY 2281 AATATATGATGATTTTATTTAACTCAGATCAAAACAGCTGTGCAACCTGCAAGAT 2340
DB 2281 AATATATGATGATTTTATTTAACTCAGATCAAAACAGCTGTGCAACCTGCAAGAT 2340
QY 2341 TTGAAGAAGTCAATTAACAGAAACATTTTGGGTAGCATCTGAAATTTCTCGAAGAAACAA 2400
DB 2341 TTGAAGAAGTCAATTAACAGAAACATTTTGGGTAGCATCTGAAATTTCTCGAAGAAACAA 2400
QY 2401 ACCACTGAAGAGGATGAAGATCATTAGCATTTTCATCAAGATAGCACTGCACTGTAGGG 2460
DB 2401 ACCACTGAAGAGGATGAAGATCATTAGCATTTTCATCAAGATAGCACTGCACTGTAGGG 2460
QY 2461 AATGCAAGAAATTTTAACTCAATGTTTGCATATCAGTGGCCTTAACCTGGCAGCAGTGG 2520
DB 2461 AATGCAAGAAATTTTAACTCAATGTTTGCATATCAGTGGCCTTAACCTGGCAGCAGTGG 2520
QY 2521 CAAGACTGGAAGCAGCTGGGGAACCTTCCCAATTAATAGCAAAACTATTTTCAGATC 2580
DB 2521 CAAGACTGGAAGCAGCTGGGGAACCTTCCCAATTAATAGCAAAACTATTTTCAGATC 2580
QY 2581 TCCAAAGACTGTTTGAATCTTCCAGAAACATGGCAAAATATGTAATGTTCTCAATAGTC 2640
DB 2581 TCCAAAGACTGTTTGAATCTTCCAGAAACATGGCAAAATATGTAATGTTCTCAATAGTC 2640
QY 2641 AAATATCAACCTCCCATTAATCCTCTATTTCCAGTTATCAAAAAGATCTCACCTTCC 2700
DB 2641 AAATATCAACCTCCCATTAATCCTCTATTTCCAGTTATCAAAAAGATCTCACCTTCC 2700
QY 2701 TTTACAGAAAGAAATGACTCAAAAGTGAAGGGGTGGTCAATTTTGAAGAGTAAAGATGA 2760
DB 2701 TTTACAGAAAGAAATGACTCAAAAGTGAAGGGGTGGTCAATTTTGAAGAGTAAAGATGA 2760
QY 2761 TTTGCAAAAGAAATTTGCTCAAGTTGGCCGAATGGCTTCACTGAACATGAGCCCTGCTTA 2820
DB 2761 TTTGCAAAAGAAATTTGCTCAAGTTGGCCGAATGGCTTCACTGAACATGAGCCCTGCTTA 2820
QY 2821 TGTTCAGGAGTCTCGGAAGAAGAAATGGCGGAGTTTGGGGTCTCTCAGCCAGGGAGTACAA 2880
DB 2821 TGTTCAGGAGTCTCGGAAGAAGAAATGGCGGAGTTTGGGGTCTCTCAGCCAGGGAGTACAA 2880
QY 2881 ATGCAACAGTGTAGATGTTGTCTCAGACAGGTGTCTATAAAAAGCGGTAAGCTGTAGTT 2940
DB 2881 ATGCAACAGTGTAGATGTTGTCTCAGACAGGTGTCTATAAAAAGCGGTAAGCTGTAGTT 2940
QY 2941 CTTTCTCAATGCCAAAAGCTTTATGAGATGCCCAATGGCTCGAAAAGTGAACAGT 3000
DB 2941 CTTTCTCAATGCCAAAAGCTTTATGAGATGCCCAATGGCTCGAAAAGTGAACAGT 3000
QY 3001 ACCTTTCCAAATTTGAGAGTGAAGATGACGAGAGAGATCTTCAAGACATTTCTCTGAGT 3060
DB 3001 ACCTTTCCAAATTTGAGAGTGAAGATGACGAGAGAGATCTTCAAGACATTTCTCTGAGT 3060
QY 3061 GTGAGCCGAGCAACCAACATTTGGCTTAAAGATCTGTGACAAAAAGCTGTCAATCCG 3120
DB 3061 GTGAGCCGAGCAACCAACATTTGGCTTAAAGATCTGTGACAAAAAGCTGTCAATCCG 3120
QY 3121 AGACCTTCCAGTACCTTCCAAAGGCGAGGTCACAACAGAAAGTCACTCCCTGTCAGAGC 3180
DB 3121 AGACCTTCCAGTACCTTCCAAAGGCGAGGTCACAACAGAAAGTCACTCCCTGTCAGAGC 3180
QY 3181 CCCAGCAGCAGCCACCACCAGACATTAATTAACACAGAGACTCAGATTTCCCGCTGT 3240
DB 3181 CCCAGCAGCAGCCACCACCAGACATTAATTAACACAGAGACTCAGATTTCCCGCTGT 3240
QY 3241 CCCTTATCTTCAGGGAAGAAAGTCCCGTAAAGCATCTCCACCTTTTGGCATAAACT 3300
DB 3241 CCCTTATCTTCAGGGAAGAAAGTCCCGTAAAGCATCTCCACCTTTTGGCATAAACT 3300
QY 3301 CTCACACAGCTTTAAAAAAATTTCTTGTCTGAAAGAAAGATTTGAGAGTCA 3360
DB 3301 CTCACACAGCTTTAAAAAAATTTCTTGTCTGAAAGAAAGATTTGAGAGTCA 3360
QY 3361 AGAAAGAGCTGAAGTATCAATCAAAATGATCTTCGAGCTTCTCTCTCTACTT 3420
DB 3361 AGAAAGAGCTGAAGTATCAATCAAAATGATCTTCGAGCTTCTCTCTCTACTT 3420
QY 3421 CTCACAGAGTCTCCAGGAAGAGCTATATCTTTGGCTCCAGTGTACTGTGATAT 3480
DB 3421 CTCACAGAGTCTCCAGGAAGAGCTATATCTTTGGCTCCAGTGTACTGTGATAT 3480
QY 3481 TTTTCAATTTCTGTCAAGAGAAATTTTTCAGATCAGTATTTGTATGCAATTCGTCT 3540
DB 3481 TTTTCAATTTCTGTCAAGAGAAATTTTTCAGATCAGTATTTGTATGCAATTCGTCT 3540
QY 3541 TTTGACTGAGTCCAGTCTCACTGACAGATGAGAGGCGCAGAGCTTCTGACATCG 3600
DB 3541 TTTGACTGAGTCCAGTCTCACTGACAGATGAGAGGCGCAGAGCTTCTGACATCG 3600
QY 3601 TCGAACAACACTTAGGATGCGAGATGAGAGGAGGAGCATGATTTGAACCTGATCACT 3660
DB 3601 TCGAACAACACTTAGGATGCGAGATGAGAGGAGGAGCATGATTTGAACCTGATCACT 3660
QY 3661 ATAGCTTGGGGTCTTATGCAACATGCTCGAGGCGGAGGCTTATGCTACAGTACAG 3720
DB 3661 ATAGCTTGGGGTCTTATGCAACATGCTCGAGGCGGAGGCTTATGCTACAGTACAG 3720
QY 3721 TAAATTTCTTCCAAAGCAGAGAGAACTTTCCAGATCAGGGGATCCGCGCTCACTTG 3780
DB 3721 TAAATTTCTTCCAAAGCAGAGAGAACTTTCCAGATCAGGGGATCCGCGCTCACTTG 3780
QY 3781 ATGCTGTGACAGTGGGCGGAGGAGTGGAGCTGACGTATCTTAATGGCTCCATGATATA 3840
DB 3781 ATGCTGTGACAGTGGGCGGAGGAGTGGAGCTGACGTATCTTAATGGCTCCATGATATA 3840
QY 3841 TACAGAGATCCAGCAGCAGAGAGAGTGGAGACTTTCCATTTGCGGCATATCTACTTG 3900
DB 3841 TACAGAGATCCAGCAGCAGAGAGAGTGGAGACTTTCCATTTGCGGCATATCTACTTG 3900
QY 3901 ATTATTCAGGGGATCTGCAAGTTATGGCATCAACAGCATATGACACCAATTTATGT 3960
DB 3901 ATTATTCAGGGGATCTGCAAGTTATGGCATCAACAGCATATGACACCAATTTATGT 3960
QY 3961 TTTTCAATCATAGCACAAGATATAACAGGCAAAATCAAAAGTAGAGAGCTTGAACAG 4020
DB 3961 TTTTCAATCATAGCACAAGATATAACAGGCAAAATCAAAAGTAGAGAGCTTGAACAG 4020
QY 4021 CCCAGTCCGAGCAAGCTGGGCGCTTTCACAGGTTACTGGGGAGAGACTCAGAGGCTG 4080
DB 4021 CCCAGTCCGAGCAAGCTGGGCGCTTTCACAGGTTACTGGGGAGAGACTCAGAGGCTG 4080

Db 4021 CCCAGTCCCGAGCAAGCTGGCGCTTCCACAGATTACTGGGGAGAAAGATCAGAAAGGTG 4080
OY 4081 ACACAGGCAAAATTAAGCGGAGGGGTGGAAAGATGTTCCATTGGAAGCCGAAGACAGTA 4140
Db 4081 ACACAGGCAAAATTAAGCGGAGGGGTGGAAAGATGTTCCATTGGAAGCCGAAGACAGTA 4140
OY 4141 GCCTAACGTCGTGTGATAGGAAAGAAACCAAGCGCTGTCCCATGCTGTGCCACATAGTGTG 4200
Db 4141 GCCTAACGTCGTGTGATAGGAAAGAAACCAAGCGCTGTCCCATGCTGTGCCACATAGTGTG 4200
OY 4201 TGGCATCAAGTACTACAAAGGGGCTTCATTGCAAGAAAGAGGGGAGTATCGAGAGCCCC 4260
Db 4201 TGGCATCAAGTACTACAAAGGGGCTTCATTGCAAGAAAGAGGGGAGTATCGAGAGCCCC 4260
OY 4261 CGGCCACCCCTCCGGGCTCATTTGGAATTCCTCATTTCTGACTTTCCAGAAAGGGGACTCCG 4320
Db 4261 CGGCCACCCCTCCGGGCTCATTTGGAATTCCTCATTTCTGACTTTCCAGAAAGGGGACTCCG 4320
OY 4321 ATCCAGCCAGAAACCGCCGAGTACAGAGTGGCCCTTCAGAGATCGCGATGGTGCAC 4380
Db 4321 ATCCAGCCAGAAACCGCCGAGTACAGAGTGGCCCTTCAGAGATCGCGATGGTGCAC 4380
OY 4381 GATCTCCGACACAGCTGGGCTTCATCCGTACAGACGCCACATCGGCGCATCCACACGA 4440
Db 4381 GATCTCCGACACAGCTGGGCTTCATCCGTACAGACGCCACATCGGCGCATCCACACGA 4440
OY 4441 GCAGGCTGTGAAACAACCTCAGTGGCATTAACCCGAGAGTGTGACCCGCGCTGGCC 4500
Db 4441 GCAGGCTGTGAAACAACCTCAGTGGCATTAACCCGAGAGTGTGACCCGCGCTGGCC 4500
OY 4501 CTTATCAGTCCCAAGGGTTTTCCACCAGAGAGATGAAGTGAACAAGTTTCTGCTGTT 4560
Db 4501 CTTATCAGTCCCAAGGGTTTTCCACCAGAGAGATGAAGTGAACAAGTTTCTGCTGTT 4560
OY 4561 GAGGCAACAGCTTTTGGAGACAGACGACCCACTGAAAGAGACACAGAGACAGT 4620
Db 4561 GAGGCAACAGCTTTTGGAGACAGACGACCCACTGAAAGAGAGACACAAAMAACGTT 4620
OY 4621 CCGGACATTTGGAGCTTGGAACTCAGATTCTGAGAGCGTGGACAGATTGCTCTTC 4680
Db 4621 CCGGACATTTGGAGCTTGGAACTCAGATTCTGAGAGCGTGGACAGATTGCTCTTC 4680
OY 4681 CCGGCTTTAAAGACAGATGGGGCTTCTCTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 4740
Db 4681 CCGGCTTTAAAGACAGATGGGGCTTCTCTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 4740
OY 4741 AATACGTGAGAAATTTGCCCTGGCACTTTTCAGACTTTGTTGCTTGAATGCAAGTGC 4800
Db 4741 AATACGTGAGAAATTTGCCCTGGCACTTTTCAGACTTTGTTGCTTGAATGCAAGTGC 4800
OY 4801 AGCAATCTTGAGCTCCCACTGTTGCTGGCTGCACATCAGACAGATATCATTTGCCAATTC 4860
Db 4801 AGCAATCTTGAGCTCCCACTGTTGCTGGCTGCACATCAGACAGATATCATTTGCCAATTC 4860
OY 4861 CAAGATCATCACAACAAGATGATTCACTGCTGCTGCACTTCTCNAATGCCGTGAAGATT 4920
Db 4861 CAAGATCATCACAACAAGATGATTCACTGCTGCTGCACTTCTCNAATGCCGTGAAGATT 4920
OY 4921 TTTTATATCTCTCTTTAGATTTCATTCAGTCTTAGACCTTGTATGCTCATTTGGGATATG 4980
Db 4921 TTTTATATCTCTCTTTAGATTTCATTCAGTCTTAGACCTTGTATGCTCATTTGGGATATG 4980
OY 4981 AGAAAGCTAGGCACTGAACTACTTGGGGCTTAAACCAACCAAGGAAGCAAGAAAA 5040
Db 4981 AGAAAGCTAGGCACTGAACTACTTGGGGCTTAAACCAACCAAGGAAGCAAGAAAA 5040
OY 5041 CAATGAATCTTTGAGTACAGTGTTCACACTTTTACATGTCCTCTTTAAAAA 5100
Db 5041 CAATGAATCTTTGAGTACAGTGTTCACACTTTTACATGTCCTCTTTAAAAA 5100
OY 5101 AAAAAATGAGTTTAAAGATTTTGTTCAGAGAGTAAATATATATCATTTAATGATTACAG 5160
Db 5101 AAAAAATGAGTTTAAAGATTTTGTTCAGAGAGTAAATATATATCATTTAATGATTACAG 5160

OY 5161 TATTTATTTAAACCTTAAGTAGGGTTGCCAGCCTGGTTCTGAAAAACCAAAATATGCCG 5220
Db 5161 TATTTATTTAAACCTTAAGTAGGGTTGCCAGCCTGGTTCTGAAAAACCAAAATATGCCG 5220
OY 5221 ACAGGCTGTGGCCACACCAAGAAGACGGGAAGACCTGGCTTGGACCTGGCTCCATG 5280
Db 5221 ACAGGCTGTGGCCACACCAAGAAGACGGGAAGACCTGGCTTGGACCTGGCTCCATG 5280
OY 5281 TCCCTTGTGCTTCACCCGGAAGTGCCTATCTCGAAGTATGAATGTTAGCCAATTAA 5340
Db 5281 TCCCTTGTGCTTCACCCGGAAGTGCCTATCTCGAAGTATGAATGTTAGCCAATTAA 5340
OY 5341 TACCAAGACACCTCATCTCTCTCTCCAGTGGATGGGGTTCTTCTGAAAACTGTGG 5400
Db 5341 TACCAAGACACCTCATCTCTCTCTCCAGTGGATGGGGTTCTTCTGAAAACTGTGG 5400
OY 5401 CACATGGCCAGGGGAGGAGTAAGTGAACCTTGTGTCTGTGAGCCTTATGAGGACAG 5460
Db 5401 CACATGGCCAGGGGAGGAGTAAGTGAACCTTGTGTCTGTGAGCCTTATGAGGACAG 5460
OY 5461 ACGGTGTATGCGGATGTCTCTCATTTGAATGATGAGCAACCCCATTTTTAA 5520
Db 5461 ACGGTGTATGCGGATGTCTCTCATTTGAATGATGAGCAACCCCATTTTTAA 5520
OY 5521 GTTATATTTCTTTGATTTTGTATTTTGAAGGTGATGGTTTGTTTTGT 5580
Db 5521 GTTATATTTCTTTGATTTTGTATTTTGAAGGTGATGGTTTGTTTTGT 5580
OY 5581 TTTTATTAAGAGAAACATTTATTAAGTGAATGATGATGCAAGTGAAGCAGTTGGATGT 5640
Db 5581 TTTTATTAAGAGAAACATTTATTAAGTGAATGATGATGCAAGTGAAGCAGTTGGATGT 5640
OY 5641 GGAAGTAAATGCCAGCTGTTTATCTGCTCTTCAAGACAGCTCCCTTATTTGAATTGGC 5700
Db 5641 GGAAGTAAATGCCAGCTGTTTATCTGCTCTTCAAGACAGCTCCCTTATTTGAATTGGC 5700
OY 5701 ATTAGGGAATTAACAAGCCTTTAAGCTGTAAAGTAAAGTCAAAAAACCGTTTAACTATGCC 5760
Db 5701 ATTAGGGAATTAACAAGCCTTTAAGCTGTAAAGTAAAGTCAAAAAACCGTTTAACTATGCC 5760
OY 5761 AGCCTTTGCAAGGAGGTTAGTCAACCAAGACATTAACCTCCAGTGGCTTATGAGAGCTG 5820
Db 5761 AGCCTTTGCAAGGAGGTTAGTCAACCAAGACATTAACCTCCAGTGGCTTATGAGAGCTG 5820
OY 5821 CATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGTATTAACCTATAT 5880
Db 5821 CATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGTATTAACCTATAT 5880
OY 5881 GACTGAATGACCCCTCCACTCATTTTGTGTTTGTGTTTGCACAGACTCCGAAAAAGTGA 5940
Db 5881 GACTGAATGACCCCTCCACTCATTTTGTGTTTGTGTTTGCACAGACTCCGAAAAAGTGA 5940
OY 5941 AGGCTCCCAATCTGAGTAGTACTCAAAATGAGAACTGCTGTCTGTGATTTTTTTC 6000
Db 5941 AGGCTCCCAATCTGAGTAGTACTCAAAATGAGAACTGCTGTCTGTGATTTTTTTC 6000
OY 6001 ATTAATTTACGCTGATCATATTGATCAGTATGAATAACGTAAATAGCTTCAATTTTAAA 6060
Db 6001 ATTAATTTACGCTGATCATATTGATCAGTATGAATAACGTAAATAGCTTCAATTTTAAA 6060
OY 6061 GGGAGATGAGTGTTTTTCACATGATCAACAAGTGTGCTGTATTAATTAATTC 6120
Db 6061 GGGAGATGAGTGTTTTTCACATGATCAACAAGTGTGCTGTATTAATTAATTC 6120
OY 6121 CTTCTGTATCATGGCAATTTGTCTACTGCTTATTAACATTTCAATATGATTTGTAAT 6180
Db 6121 CTTCTGTATCATGGCAATTTGTCTACTGCTTATTAACATTTCAATATGATTTGTAAT 6180
OY 6181 TTACATGTAATATGATTAATTTTCCAGTTTATTAATATAGGCTATGAGCTCAT 6240
Db 6181 TTACATGTAATATGATTAATTTTCCAGTTTATTAATATAGGCTATGAGCTCAT 6240

```
OY 6241 ATGANAAGACAGAAATCTAGCTCTACACAACTGTCACAAATGTTATCTGAAGCATTAGT 6300
      |||||||
Db 6241 ATGANAAGACAGAAATCTAGCTCTACACAACTGTCACAAATGTTATCTGAAGCATTAGT 6300
OY 6301 AATTGTGAACATAGAGACGTATATCTCAGTTCGCTCTGTGATGTCAGAGTGAATGA 6360
      |||||||
Db 6301 AATTGTGAACATAGAGACGTATATCTCAGTTCGCTCTGTGATGTCAGAGTGAATGA 6360
OY 6361 CAATTAAGTGTGATTTCTCATATCTTTGATACACTGTTACCTGTATGCTTTTAGAA 6420
      |||||||
Db 6361 CAATTAAGTGTGATTTCTCATATCTTTGATACACTGTTACCTGTATGCTTTTAGAA 6420
OY 6421 AGACATTGTGAGAGTCTGTATCCCTTTTGTATTTTAATACATATTTAGATTTGCT 6480
      |||||||
Db 6421 AGACATTGTGAGAGTCTGTATCCCTTTTGTATTTTAATACATATTTAGATTTGCT 6480
OY 6481 TATATTTTGTGAAGATGTGAGAAATGTACTATGTTTATGCTTCTACATCCAGTTTGA 6540
      |||||||
Db 6481 TATATTTTGTGAAGATGTGAGAAATGTACTATGTTTATGCTTCTACATCCAGTTTGA 6540
OY 6541 CAAGCTGGAATAATTAATATATACAT 6568
      |||||||
Db 6541 CAAGCTGGAATAATTAATATATACAT 6568

RESULT 3
US-09-728-952-68
; Sequence 68, Application US/09728952
; Patent No. US2002011302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha T.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US2002011302A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 799
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PL_genes Version 2.0
; SEQ ID NO 68
; LENGTH: 6383
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70)..(5025)
US-09-728-952-68

Query Match 24.2%; Score 1587; DB 10; Length 6383;
Best Local Similarity 66.0%; Pred. No. 0;
Matches 2529; Conservative 0; Mismatches 1205; Indels 99; Gaps 12;
```

```
Db 817 GAAAGCAGCGATTATCTGATATCTATCAGCTTACGAGCTACGAGAGAGTAGAGATGTAGAT 876
      |||||||
OY 285 CTGAGTGGTGTGCGAGAAACAGCAGTGGATTCCGAAGACGACGATGATGACAGACATT 344
      |||||||
Db 877 TTGACAGCTCTCCAGAGAGACCTGTTGATTGTAGAGATGACGAGAAAGATGAAG 936
      |||||||
OY 345 GAGAGACATCATCTCTCTGATGAGAGGACATGTGAGACACTGCTAGAGAGAGC 404
      |||||||
Db 937 ATTATGCAACAGATCCATTGTCAGGGGGCAGATCTGTGTGAGAAATGCTTGTAAAAAGAA 996
      |||||||
OY 405 CCATTTGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 464
      |||||||
Db 997 CCTGACAGCAAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1056
      |||||||
OY 465 GCTTTTGGCAATATGACAAATGTCAGTGAAGCGGAACTGCTGCTGATGATGATGATGATGAT 524
      |||||||
Db 1057 GCATTTGCAAAACATGACCCATGCTGTGAAGAGAGAGAACTGCTGCTGATGATGATGATGATGAT 1116
      |||||||
OY 525 GTGGTGAAGAGCAGGAGCAGGACCATGATGTTAAATGATGATGATGATGATGATGATGATGAT 584
      |||||||
Db 1117 GTGTAGAGCAGCGCTGAGACTATTAATCTTGAAGATGGGCAAGAGCTTGAATCTGATAT 1176
      |||||||
OY 585 GTGATTTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 644
      |||||||
Db 1177 GTTATTTTAAAGCGCAGCTGTGAAATCAGTCAATGATGATGATGATGATGATGATGATGATGAT 1236
      |||||||
OY 645 ATGGGAATAGTTTGGTGTCTCTCTACATGACCAAGATGATGATGATGATGATGATGATGATGAT 704
      |||||||
Db 1237 ATGGGAATAGTTTGGTGTCTCTCTACATGACCAAGATGATGATGATGATGATGATGATGATGATGAT 1296
      |||||||
OY 705 AGAACAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 764
      |||||||
Db 1297 AGGACTAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1336
      |||||||
OY 765 CTGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 824
      |||||||
Db 1357 TTAACCATGTGAAATAATATACCATTAAGTTGAGAAAGAGAGGAAATGTTATGGTA 1416
      |||||||
OY 825 AAGAACACCGAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 884
      |||||||
Db 1417 CATGACATCGGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1476
      |||||||
OY 885 ACCTGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 944
      |||||||
Db 1477 ACACCTGAGCGTGCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1536
      |||||||
OY 945 ATGAAGACTTCTGTTGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1004
      |||||||
Db 1537 ATGAAGACTTCTGTTGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1596
      |||||||
OY 1005 AAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1064
      |||||||
Db 1597 AACTATTTGAAATGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1656
      |||||||
OY 1065 TTGTGGTAAATATCACTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1124
      |||||||
Db 1657 TTATGGTAAATATCACTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1716
      |||||||
OY 1125 GAAGAATTTGAANAATCTGGAAGAGAGAAATGGGTGACACCTGATGATGATGATGATGATGATGATGATGAT 1184
      |||||||
Db 1717 GAGGAATTTGAANAATCTGGAAGAGAGAAATGGGTGACACCTGATGATGATGATGATGATGATGATGATGAT 1776
      |||||||
OY 1185 ATCGGTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1244
      |||||||
Db 1777 ATCGGTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1836
      |||||||
OY 1245 GCTCCTTTGCTTTTATCTTACTGAGAGCTGTGAAGAGGATTTGGAATCTTGTGAC 1304
      |||||||
Db 1837 TCCCTTCAATCAATTCAGCTTAATGAGAGGAGTGAAGAGGATTTGATATTTTGTGAA 1896
      |||||||
OY 1305 ACTGTAGATTCAGTGAAGAACCAATGAAACAGGCTTGAAGAGGAGATCAATATTA 1364
      |||||||
```

Db 1897 GGAGTGAACCTGTGATGCAAAAGCTGCTGATTCAGACTGAAACGTGTGATCAGATTATG 1956
QY 1365 GAAGTAAATGGCCAAACCTTTGAAAACATTCAGCTGTCAAAACCTATGAAATTCCTAGA 1424
Db 1957 GAAGTAAATGGCAAAACCTTTGAAATATTTATGAAAGCCGCTTGAAATTTTGAGG 2016
QY 1425 AATAACACATCTATCTATCTAGTGAAGAAACATTTATTTATTTAAAGAACTCTTA 1484
Db 2017 AATAATACATCTATCTAGTGAAGAAACATTTATTTATTTAAAGAACTCTTA 2076
QY 1485 ACAAGATTTGTCAAGAGAAAAGAAATGCTGCCCCACCTCTCTAAATTTGTGACATT 1544
Db 2077 TTTAG---GACTGAACAGAGAAATCTGTGTCTCATATTCCTCAAAATTTGCTGAAAA 2133
QY 1545 AAAAAGGCCGTGCTACTCCATTCATCTGCTGTGTATGATAGAACAGGATGATGA 1604
Db 2134 AAAAGT---AATGCCATTTCTATTCAGCATGTGCCAGAGATATTTGAAACA---GACATCA 2187
QY 1605 CTGAAAAAGTGAACAAAAAAGTAAAGCCAGACCTGTG---GGAGAGAGAAACAGCTG 1661
Db 2188 CAGAGAGAAAGAGTAAAGAAAGTTAAAGCAATATCTGTTTCAGCTGCAAGAAACAAATC 2247
QY 1662 AAAAGATCTGACAAAGACTCGATCATGTCTGACACAGAAACCATCAATGATATT 1721
Db 2248 AGGAAGATTTGGATAAACACGATTTAGTATCTTGCTCCAAAGCTATTAGATGA 2307
QY 1722 GGGATTTGCTAGTCTCAAGATGACAGATATGATGATTAAGGACAGAAAGCATCCCA 1781
Db 2308 GGGCTAAAGCAATCAACAGATGACAGCATTTGTGGAAACAGGCACTAGACATGCTG 2367
QY 1782 ACTGATTTGCTGCTGAGTGAACCTTATCATCCAGTAAATCTGATTTATTTAGCTGAT 1841
Db 2368 GCTATATATGCCCATCCCTGGAACACATCTCATCCAGACGCCCTGATCTCTGCAACCTACC 2427
QY 1842 CATGCAATTTTAAGCTCACTGCTACTCTGACTTGGCAGATCAAGTCTAAGGGTTTTT 1901
Db 2428 ACCAGATTTGTGATTTTCCAAATCTCTTCAGATATCCCTGATCAAGTTATTAAGGTTTTC 2487
QY 1902 AAGGCTGATGAGAAAGCCCTACATCATGATCAGTAAAGACACTACAGAAAGAAAGTG 1961
Db 2488 AAGGTGATGAGAAAGTCTACATTTATCATGATTAAGAACACACAGCTAAAGAAAGTA 2547
QY 1962 GTCATGAGCTATACAGGAGGTTTTGCTGTTACTGCCCCGATCAATTTACTATCTATG 2021
Db 2548 GTTTTTCATGCTGTTCATGATTTGGTTTGACCGGTGCATCCGACATATTTCTCTGT 2607
QY 2022 GAGGTCTCTGTCAACCTGAGGAGTAAATCAACAAAGAAAGCTTCAGATCAGCTTTC 2081
Db 2608 GAAGTTTCTGTACTCTGTAGGGGTGTCATAAAGAGAAAGACTTCCAGATCAAGTTCTCC 2667
QY 2082 AAATCTGCAGACAGATATCAACTGAGTGAAGGATATATCTGAAAAACAATGGAACA 2141
Db 2668 AAATTTACTGATGAATTAACATCAATGAGAGGATATTAATAAATAACATGGAACA 2727
QY 2142 GAAACCTTTTTCAGATGAAGATGCTCAGAGTTGTTGAGAGAGATCAATTTCCCTC 2201
Db 2728 GAAACCTTTATGTTTCAAGTGAAGATGCTCAAGAACTAGTTAAGAAAGCCAGCTATCCATG 2787
QY 2202 CTTCAGCTCAGACTGTGAAGTTGCAACAGAGCTCTCTATGCGAAATTTTGAACCTTT 2261
Db 2788 CTGAGCTCAGTACATTTGAGTGCGCCAGCTGTCATGAGGAGCTTTGATTTT 2847
QY 2262 CGCAACCTTAACCTTACTGATATATATGATGATTTATTTAACTCAGATCAAAACAGC 2321
Db 2848 CGTAATTTTGAACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2907
QY 2322 TGTGCCAACCTGAAGATTTGAAGAGTCAATTAACAGGAAACATTTTGGTAGATCT 2381
Db 2908 AATATCTATTTGAAGAGTTTGAAGAGATTTGAAGAGAGATTTCTGGTGGCTCA 2967
QY 2382 GAAATTTCTAGAGAAACAAACAGCTGAAGAGATGAAGATCAATTAAGCATTTTCATCAG 2441
Db 2968 GAAATTTTAACTGAACGAATCAAGCTCAACGAATGAAGATTTATTAAGCATTTTATTA 3027

QY 2442 ATAGCACTGACCTGTAGGGAATGCAAGAAATTTTAACTCAATGTTTGCATATCATCAGTGGC 2501
Db 3028 ATTGCATCTTATTTGAGAGATTAAGAACTTCAATTTCAATTTGAAATTAAGTGGC 3087
QY 2502 CTAAACCTGCGACAGTGGCAAGACCTGCAAGACACCTGGGAGAAACCTCCCAATTAATAC 2561
Db 3088 TTGACCTGTCATCTGTAGACAGATCTGAGAAAGCTTTGGGAAAAAGTTTCCAGCAATATAC 3147
QY 2562 GAAAACTATTTCAAGATCTTCAAGACCTGTTGATCTTTTCAGAAACATGCAAAATAT 2621
Db 3148 GAGAAACATCTCAAGATCTTCAAGACATTTTGTATCATCTAGAAACATGCAAAATAT 3207
QY 2622 CGTAATGTTCTCAATGTCAAATCTCAACACCTCCCATATCCCTATTTCCAGTATAC 2681
Db 3208 AGAATATTTCTTATGATCTCAAGATGACAGCTTCATATTTTCCATCTTCCGTTGTC 3267
QY 2682 AAAAAGATCTCACCTCTCTTCAAGAGAAATGACCAAAAGTAGACGGGCTGTCAAT 2741
Db 3268 AAGAAAGATATGACATTTTCTCATGAAAGAAATGACTCCAAAGTATGATGTTAGTAAAC 3327
QY 2742 TTTGAGAGCTTAAGATGATTTGCAAAAGAAATTCGTCAGCTTGGCCGATGCTTCACTG 2801
Db 3328 TTTGAGAAAGTTAAGATGATTTTCAAGAAATCCGCCAAGTTGTTGATGATGATGATGATG 3387
QY 2802 AATATGACCTGCTCCATGTTCTGAGACCTCGAAGAAAGAAATGCGGATTTGGGCTCT 2861
Db 3388 AATATGACCTCCATGATGATTTT-----TCCACAGAGTCA 3423
QY 2862 CTCAGCAGGATGATCAAAATGCAACAGTGTAGATTTGCTGAGAGGCTGATCAAA 2921
Db 3424 CTGAGTAAAGAAAGCAAAATTTCAACATGCTGATTT---TCAGGAGGCTGCTACAAA 3480
QY 2922 AAGCGGTACGTGTGATCTCTTCTCAATGCCAAAAAGCTTTATGAAGATGCCCAAT 2981
Db 3481 AAAAGGCGACGCGCAGCTCTCTCTTAATGCGAAAGCATATGATGATGCCCAATG 3540
QY 2982 GCTCGAAAGTGAAGCATGATCTTCCATTTGAGCTAGAAATGAGAGAGATCTT 3041
Db 3541 GCAAGAGAGTGAAGCATGATCTTCTTCCAGTCTGATGATGAGAGAGAGAAAGTTTC 3600
QY 3042 CAGCATTTATCTGTGAGTGTGAGCCAGCAACCAACATTTGCTTAAGAAATCCGTTGAC 3101
Db 3601 CAGATGATGATTAATGATGAGGAGCTGATGATGATGATGATGATGATGATGATGATG 3660
QY 3102 AAAAAGCTGTCAA---ATCCGAGACTTTCAGTATCTCCAAAGGAGGCTTCAACAAG 3158
Db 3661 AAAAGATCAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3720
QY 3159 AAGGCTAGTCCCTGCGCAGACGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3218
Db 3721 ACTAAAGCCCATTTGCTATCAACC-----CAGAGGTAAAGCTAG 3759
QY 3219 GAGCTAGAGTTCGCCGCGCTTATCTTATCTTACAGGAAAGAGGCCGTAAAGAT 3278
Db 3760 GTGCTTAGGTGCGCAGCTGTTAATTTTCAACCCATCAAGAGAGAGGACAAACAAAGAC 3819
QY 3279 CTCACACCTTTTGGCATTAACCTTCCACAAGCTTTTAAAAAAAATTTCTTTTGTGAA 3338
Db 3820 CTGACATGAAATCAAAAGTTTACGTCAAGAAAGTTTAAAGAACAA-----CTGAA 3887
QY 3339 GAAGGAAGTTTGAAGCTCACAGAAACAGGCTGAAGATATCAATATGATGATCTTCG 3398
Db 3868 GAAATTA-----GTGTAAGAACATTAACAGAACATATTTCTGTGGGCTCATCT 3918
QY 3399 CAGTTTCTTCT 3458
Db 3919 TTACATTTCTAGTCT 3978
QY 3459 CCAAGTGTACTGTGATTAATTTTTCAGATTTCTGATGATGATGATGATGATGATGATG 3518
Db 3979 CCAATCAGCTAAATGATGACAACTGCTGACCTCCAGCATATGATGATGATGATGATGATG 4038

```

OY 3519 AGATTTTACCAATTCCTTTTGAAGTCCAGTCTCAGTGCAGATGAGAGCCG 3578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4039 AGCATTCGTAGCAATTTCTTTGTTGATCTCATCTCTGACGCTTACAGATGAGAGGCTG 4098
OY 3579 CAGAGGATTCCTGACATCTGTGAAACCAACCTAGGGATGGGAGATGAGAGCCG 3638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4099 TCCTCTCAGGCGCTTGGAGATCCCTGAAATCCACTGGGGCATTTGGAAAAGACAGCGT 4158
OY 3639 ACCATTTGAACCTGATCAGTATAGCTTGGGGTCCATGCACCAATGTCCAGGGCCGA 3698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4159 TCAGGATAGGAGATCATATGCAACATGCGCTGGGTGACACTTTGAAGCCATCTCTA 4218
OY 3699 GCGTTATATGCTACAGCTCAGTATTTCTCTCCAGCAGAGAACTTCCAGAGT 3758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4219 ATCAAGTTTAACTGTCTCA-----TCGTCTGTGAGCATGAAAGATTTCTCAAGAG 4272
OY 3759 CAGGGGATCGCGGCTCAGTCTGATGCTGCTGACAGTGGCGCTGGAGCTGAGCTCATGC 3818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4273 CATATCATTT-----ATGAAGCAGCTGACAGTGTCTGTGAAAGTTGGACTTGGCTGT 4323
OY 3819 TCAAGTGGCTCCCATGATATATATACAGAGATCCAGCAGAGAGAGCTGGGA 3871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4324 TCAAGCAGCTCCCATGACAACTTCCAAAGCCTTCCAAACCAAAAGAGCTGGGA 4376

RESULT 4
US-10-046-935-1879/c
; Sequence 1879, Application US/10046935
; Patent No. US2002015601A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1879
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-046-935-1879

Query Match          6.1%; Score 399; DB 9; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.8e-81;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5961 ACTCAATGTGAGAGACTGCTGCTGATTTTTCATTAATTCAGCTGATCATTA 6020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 ACTCAATGTGAGAGACTGCTGCTGATTTTTCATTAATTCAGCTGATCATTA 340
OY 6021 TTGATCGATGATTAACGTAATAGCTTCAAAATTTTAAAGTGAATGCAATGTTTTT 6080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 TTGATCGATGATTAACGTAATAGCTTCAAAATTTTAAAGTGAATGCAATGTTTTT 280
OY 6081 CACTGTATCAACAATGTCACTGCTTATTTAATATTCCTCTGTATATATGAGCAATTTG 6140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 CACTGTATCAACAATGTCACTGCTTATTTAATATTCCTCTGTATATATGAGCAATTTG 220
OY 6141 TCTACTTCTTATTAACATTTGATTTGATTTTGTATTTTACATTAATATGATCATTA 6200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 TCTACTTCTTATTAACATTTGATTTGATTTTGTATTTTACATTAATATGATCATTA 160
OY 6201 TTGCGAGTTTATTTATAGCTATGAGCTCATGTCATATAGAAAGACAGAAATCTAG 6260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TTGCGAGTTTATTTATAGCTATGAGCTCATGTCATATAGAAAGACAGAAATCTAG 100
OY 6261 CTCTACCAACAAGTTGCACAATGTATCTTAAGCATTAATTTGTAGAACATGAGACTG 6320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db 99 CTCTACCAACAAGTTGCACAATGTATCTTAAGCATTAATTTGTAGAACATGAGACTG 40
OY 6321 CTAACTCAGTTCCCTCTGTGATGTCAAGTGCAGATGT 6359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39 CTAACTCAGTTCCCTCTGTGATGTCAAGTGCAGAAAGT 1

RESULT 5
US-09-878-178-1879/c
; Sequence 1879, Application US/09878178
; Patent No. US2002017552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1879
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-878-178-1879

Query Match          6.1%; Score 399; DB 9; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.8e-81;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5961 ACTCAATGTGAGAGACTGCTGCTGATTTTTCATTAATTCAGCTGATCATTA 6020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 ACTCAATGTGAGAGACTGCTGCTGATTTTTCATTAATTCAGCTGATCATTA 340
OY 6021 TTGATCGATGATTAACGTAATAGCTTCAAAATTTTAAAGTGAATGCAATGTTTTT 6080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 TTGATCGATGATTAACGTAATAGCTTCAAAATTTTAAAGTGAATGCAATGTTTTT 280
OY 6081 CACTGTATCAACAATGTCACTGCTTATTTAATATTCCTCTGTATATATGAGCAATTTG 6140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 CACTGTATCAACAATGTCACTGCTTATTTAATATTCCTCTGTATATATGAGCAATTTG 220
OY 6141 TCTACTTCTTATTAACATTTGATTTGATTTTGTATTTTTCATTAATATGATCATTA 6200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 TCTACTTCTTATTAACATTTGATTTGATTTTGTATTTTTCATTAATATGATCATTA 160
OY 6201 TTGCGAGTTTATTTATAGCTATGAGCTCATGTCATATAGAAAGACAGAAATCTAG 6260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TTGCGAGTTTATTTATAGCTATGAGCTCATGTCATATAGAAAGACAGAAATCTAG 100
OY 6261 CTCTACCAACAAGTTGCACAATGTATCTTAAGCATTAATTTGTAGAACATGAGACTG 6320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 CTCTACCAACAAGTTGCACAATGTATCTTAAGCATTAATTTGTAGAACATGAGACTG 40
OY 6321 CTAACTCAGTTCCCTCTGTGATGTCAAGTGCAGATGT 6359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39 CTAACTCAGTTCCCTCTGTGATGTCAAGTGCAGAAAGT 1

RESULT 6
US-09-911-826a-3
; Sequence 3, Application US/09911826A
; Patent No. US2002014316A1
; GENERAL INFORMATION:
; APPLICANT: Rolin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: DDM-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
; CURRENT FILING DATE: 2002-02-26
```



```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-878-178-1196
Query Match
Best Local Similarity 99.8%; Score 388; DB 9; Length 400;
Matches 399; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 5961 ACTCAATGTCAGAACTGCTGCTGATTTTTCATTAATTCAGTGCATCA 6020
    |||||||
DB 400 ACTCAATGTCAGAACTGCTGCTGATTTTTCATTAATTCAGTGCATCA 341
    |||||||

OY 6021 TTGATCAGTAGATAACGTAATAGCTTCAATTTTAAAGTGCAATGCACTG-1TTT 6079
    |||||||
DB 340 TTGATCAGTAGATAACGTAATAGCTTCAATTTTAAAGTGCAATGCACTG-1TTT 281
    |||||||

OY 6080 TCACAGTATCAACAATGTCAGTCTTATTTATTAATTCCTTCATCAGGCACTT 6139
    |||||||
DB 280 TCACAGTATCAACAATGTCAGTCTTATTTATTAATTCCTTCATCAGGCACTT 221
    |||||||

OY 6140 GTCACCTGCTTATTCATGTCATTTGTAATTTTACATGTAATGCACTTA 6199
    |||||||
DB 220 GTCACCTGCTTATTCATGTCATTTGTAATTTTACATGTAATGCACTTA 161
    |||||||

OY 6200 TTTGCCAGTTTATTTATAGCTATGAGCCTCATGTCATATGAAAGACAGAACTCA 6259
    |||||||
DB 160 TTTGCCAGTTTATTTATAGCTATGAGCCTCATGTCATATGAAAGACAGAACTCA 101
    |||||||

OY 6260 GCTCACCACCAAGTGCACCAATGTTATTAAGCATTTGAGACATAGGACT 6319
    |||||||
DB 100 GCTCACCACCAAGTGCACCAATGTTATTAAGCATTTGAGACATAGGACT 41
    |||||||

OY 6320 GCTAATCTCAGTTCGCTCTGTGATGTCAGAGCAGAACTG 6359
    |||||||
DB 40 GCTAATCTCAGTTCGCTCTGTGATGTCAGAGCAGAACTG 1
    |||||||

RESULT 9
US-09-733-607-15/C
; Sequence 15, Application US/09733607
; Patent No. US20020042054A1
; GENERAL INFORMATION:
; APPLICANT: Tononi, Giulio
; APPLICANT: Cirelli, Chiara
; APPLICANT: Shaw, Paul J.
; APPLICANT: Greenspan, Ralph J.
; TITLE OF INVENTION: Vigilance Nucleic Acids and Related
; FILE REFERENCE: P-NI 4447
; CURRENT APPLICATION NUMBER: US/09/733,607
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US 09/456,785
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 263
; TYPE: DNA
; ORGANISM: raltus
US-09-733-607-15
Query Match
Best Local Similarity 91.6%; Score 227.8; DB 10; Length 263;
Matches 241; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 526 TGGTGAAGAAGCAGGACCATAGTGTAAATGATGCTGAGAGCTGCTGCTGAG 585
    |||||||
DB 263 TGGTGAAGAAGCAGGACCATAGTGTAAATGATGCTGAGAGCTGCTGCTGAG 204
    |||||||

OY 586 TGATTCATGATCTGCTGAGAGTGCATATCCAGATGGAAGCAAGAACTCTGCA 645
    |||||||
DB 203 TGATTCATGATCTGCTGAGAGTGCATATCCAGATGGAAGCAAGAACTCTGCA 144
    |||||||
```

```
OY 646 TGGCAATAGTTTGTCTCTCTCCATCCATGACCAAGAAATACATGAAGAGTATGA 705
    |||||||
DB 143 TGGCAAGAGCTTTGGTTTGGCCACCATGACCAAGAGTATCAAGAGTCAATGA 84
    |||||||

OY 706 GAACAAAGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765
    |||||||
DB 83 GGACAAAGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24
    |||||||

OY 766 TCATATCAAGTAGAAAAAACAATG 788
    |||||||
DB 23 TGACCAAGTAGAAAAAACAATG 1
    |||||||

RESULT 10
US-09-864-761-27564/C
; Sequence 27564, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DEIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27564
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004622.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.65
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.6
```


PRIOR APPLICATION NUMBER: US 60/180,312

OY 2257 TCTTTCGCAACATGTGAACCTACTGATATATAGATGATTATTTTAAACTCGATCAAAAA 2316
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 405 TGTTTTCGTAATAATGTGAACCCGACTGAGTAGCATCGAAGACCCTTTTAAAGTTAAATTCAAAA 346
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 OY 2317 CCAGCTGTGCCAACCTGAAGAGATTGSAAGAAGTCATTAAACGAGAAACATTTGGCGTAG 2376
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 345 CAGGAAATATCATTGTTGAAGAGGTTTGAGGCATTTGTAAACCAAGACATTTGGGTTG 286
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 OY 2377 CATCTGAATTTCTCGAAGAAACAACCCAGCTGAAGAGAGATGAAGATCATTAAAGCATTTCA 2436
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 285 CCTCGAATTTTAACTGAAGCAATCATGCTCAACGAATGAAGTATTTAAAGCATTTTA 226
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 OY 2437 TCACATATGACACTGCACGTGTAGGCAATGCAGAANTTTAACTCAATGTTTCATATCA 2496
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 225 TTAAATATTCACCTTCATTGTGAGATGTAAAGACTTCAATTCATGTTTCGAATAATTA 166
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 OY 2497 G 2497
 |
 Db 165 G 165

RESULT 12
 US-09-864-761-5303/c
 : Sequence 5303, Application US/09864761
 : Patent No. US20020048763A1
 : GENERAL INFORMATION:
 : APPLICANT: Penn, Sharon G.
 : APPLICANT: Rank, David R.
 : APPLICANT: Hanzel, David K.
 : APPLICANT: Chen, Wensheng
 : TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 : TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 : FILE REFERENCE: Aegmica-X-1
 : CURRENT APPLICATION NUMBER: US/09/864,761
 : CURRENT FILING DATE: 2001-05-23
 : PRIOR APPLICATION NUMBER: US 60/180,312

```

PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 5303
LENGTH: 471
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004227.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
US-09-864-761-5303
```

```

Query Match 1.6% Score 104.6; DB 10; Length 471;
Best Local Similarity 69.1%; Pred. No. 2.5e-14;
Matches 143; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
```

```

QY 1155 AAATGGGTGGACCTAAGCCTGTTGATATCGCGTGTGCTGCTTAAGCAAAAGAGA 1214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 471 AAGATGAATGTCATCTCGGTATGATATGCTGTGCGCAAGGCTAAGTGAGGA 412
QY 1215 TTGATGACGTAAACAACATCCGAGAAGCTCTTTCCTTTATCTTACTTGGAGC 1274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 411 CAGGTGTGTGTCMAAAGGCTTCCCGAGTCCCTCTCAATTCAGCCTTAATGAGG 352
QY 1275 TCTGAGAAGGATTTGGAATCTTTGTTGACAGTGTAGATTGAGTAGCAACAACCTGA 1334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 351 AATGAGAAGGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 292
QY 1335 GCAGGCTGAACGGGGGATCAGATA 1361
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 291 TCAGGACTGAACGTGTGTATCAGTA 265
```

```

RESULT 13
US-09-954-456-1970
; Sequence 1970, Application US/09954456
; Patent No. US20020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
CURRENT FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1970
LENGTH: 5900
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-1970
```

```

Query Match 1.6% Score 104.6; DB 10; Length 5900;
Best Local Similarity 55.2%; Pred. No. 9.3e-14;
Matches 228; Conservative 0; Mismatches 179; Indels 6; Gaps 1;
```

```

QY 2355 AACGGAACATTGTTGGTAGCATGTAATTCAGAGAAACAACCAAGCTGAAGAG 2414
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1235 AATGAGTCCAGCTTTGGGTGGCGACGAGATTCGTCTGCAGCCAGCTGGGCAAGCA 1294
QY 2415 ATGAAGATCATTAAGCATTTTCATCAGATAGCACGTGCTGAGGGAATGCCAATTTT 2474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1295 GTCCAGCTGGTGAATAAATTCATCAATATGCGGCTCAGCAAAAGCCCAAGAAACCTG 1354
QY 2475 AACATCATGTTGGATCATCATCAGTGGCTTAACCTGACACCACTGGCAAGCTGCAAG 2534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1355 AATCTTTCTTTGTCATGATGATGGGTCTCAACACCTCTCTGATGATGATGATG 1414
QY 2535 ACCTGGAGAAACTTCCCAATTAATACGAAAACTATTTAAGATCTCCAAAGCTGTTT 2594
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1415 ACCTGGAGAAATCCCTGGGAAGTTAAGAACTTTCTCTGAACTGTAAGATTTAACA 1474
QY 2595 GATCCTTCAGAAACATGCGCAAAATATCTGAATGTTCTCATATGTCAAATATCTACA 2654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1475 GATCCTTCCTTAATTCACAAAGCCTTAAGATGATCATTTCA-----AAAGATGAAGCCA 1528
QY 2655 CCCAATATCCCTATATCCAGTATATCAAAAGAGATCTCACTCTTCAAGAGAAAT 2714
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1529 CCAAAATCCCTTCATGCTTATTTGCTTAAATGATTAATTAATTAATTAATTAATTA 1588
QY 2715 GACTCAAAAGTACAGGGGCTGCAATTTTGAAGAGCTTAAGGATGATTGCAAA 2767
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1589 AAAACTTTTGGATTAATCTTGAATTTGAAAGAGCTGATATGATGAGCA 1641
```

```
RESULT 14
US-09-822-846-129/c
; Sequence 129, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallee, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steinger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fehchel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 129
; LENGTH: 2536
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-129

Query Match      1.5%; Score 95.8; DB 9; Length 2536;
Best Local Similarity 52.6%; Pred. No. 6.1e-12;
Matches 236; Conservative 0; Mismatches 207; Indels 6; Gaps 1;

QY 2325 GCCACCTGAGAGATTGGAAGATCATTAACGAGAACTTTGGTAGCATCGAA 2384
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2527 GCCACTGGAGCTGCTGCGTCGAGCGCTGAGGAGGTACGACACTGGCGCCGAA 2468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2385 ATTCTCAGAAACAAACAGCTGAGAGAGATGAATCAATTAAGCATTCATCAAGATA 2444
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2467 GTCTGCTGTGCGAGCGCCGCGCAACGCCGCGAGCTGCTCAAGAAATTCATCAAGATC 2408
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2445 GCACGTGACGTGAGGAGATGCAAGAAATTTTAATCAATGTTGCAATCATCAGTGCCTA 2504
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2407 GCGGCGCTGCGAAGCAAGAGAGACTCTCTCTTCTTACGCCGCGGTCATGCGGCTG 2348
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2505 AACCTGACACGAGTGGCAAGACCTGTTGATCTTCCAGAAACATGCAAAATATCGT 2564
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2347 GACAAGCGCCGTGTGACGCCCTTCCACTGAGGAAAGCTGCCAGGAAATTCGAAG 2288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2565 AAATATTTCAGATCTCCAAAGCTGTTGATCTTCCAGAAACATGCAAAATATCGT 2624
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2287 AACTGTTTGGCAAAATTTGAGAACTGAGGAGCCCTCGAGAAACCAAAAGCTACGA 2228
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2625 AATGTTCTCAATAGTCAAAATCTACAACCTCCCATTAATCCCTCATTCAGATTACAA 2684
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2227 GAAGTGATC-----TCCAAATGAACCCCTGTGATTTCCCTTCGCTCATGATCTC 2174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2685 AAGGATCTACCTCTCTTCCAGGAAGAAATGACTCAAAATAGCGGCGTGGTCAATTTT 2744
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2173 AAAGACCTGACTTTCTGCGCAAGAGGAGTAAGCCCTTGTAGTGTGTTGTAACATC 2114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2745 GAGAAGCTAAGATGATTTGCAAAAGAAAT 2773
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2113 GAGAAGCTGATTCAGTGGCCGCAAAAGT 2085
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 15
US-09-864-761-22078/c
; Sequence 22078, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemlica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22078
; LENGTH: 182
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004227.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
; OTHER INFORMATION: SWISSPROT HIT: Q13796, EVALU1 1.30e+00
; OTHER INFORMATION: NT HIT: q17706512, EVALU2 2.00e-98
; OTHER INFORMATION: EST_HUMAN HIT: BE552426.1, EVALU6 6.30e-01
US-09-864-761-22078
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 10:34:29 : Search time 203 Seconds

(without alignments)
922.428 Million cell updates/sec

Title: US-09-911-826a-1

Perfect score: 6568
Sequence: 1 ctgcgcacgtagagatg.....aaaataataataacat 6568

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued_Patents_NA:*

- 1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/CTCUS.COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	62.4	1.0	7218	1	US-08-232-463-14
C 2	60.8	0.9	5398	3	US-09-356-952-11
C 3	58.2	0.9	43676	3	US-09-356-952-12
C 4	51.6	0.8	7218	1	US-08-232-463-14
C 5	51.6	0.8	19124	2	US-08-487-8268-13
C 6	48	0.7	3275	4	US-09-370-838-151
C 7	47	0.7	1474	4	US-08-821-994-64
C 8	47	0.7	6200	4	US-09-439-923-1
C 9	45.8	0.7	2550	6	5258287-23
C 10	45	0.7	2805	4	US-09-653-839-9
C 11	44.8	0.7	1927	4	US-09-336-536-66
C 12	44.6	0.7	140	1	US-08-628-417-5
C 13	44.6	0.7	240	1	US-08-628-417-6
C 14	44.4	0.7	1582	3	US-08-545-1968-10
C 15	44.4	0.7	1582	3	US-08-545-1968-12
C 16	44.4	0.7	2246	4	US-09-363-708-3
C 17	44.2	0.7	1046	1	US-08-361-467B-4
C 18	44.2	0.7	1046	1	US-08-484-332C-4
C 19	44.2	0.7	3268	3	US-09-356-952-13
C 20	44.2	0.7	1508	4	US-09-039-046-1
C 21	43.8	0.7	1279	3	US-09-248-335-25
C 22	43.8	0.7	2665	4	US-08-971-089-5
C 23	43.8	0.7	3410	4	US-09-020-956-110
C 24	43.8	0.7	3410	4	US-09-030-607-110
C 25	43.8	0.7	3410	4	US-09-605-785-110
C 26	43.8	0.7	3410	4	US-09-439-313-110
C 27	43.8	0.7	3410	4	US-09-352-616A-110

C 28	43.8	0.7	3410	4	US-09-602-877A-100	Sequence 100, App
C 29	43.8	0.7	3410	4	US-09-232-149A-110	Sequence 110, App
C 30	43.6	0.7	1027	4	US-09-465-558-57	Sequence 57, App1
C 31	43.6	0.7	1098	3	US-09-248-335-35	Sequence 35, App1
C 32	43.6	0.7	1872	4	US-09-801-052-1	Sequence 1, App11
C 33	43.4	0.7	144	1	US-08-702-344-26	Sequence 26, App1
C 34	42.8	0.7	1100	4	US-07-861-458C-4	Sequence 4, App11
C 35	42.8	0.7	1117	4	US-09-247-373B-33	Sequence 33, App1
C 36	42.6	0.6	1069	4	US-09-372-422A-7	Sequence 7, App11
C 37	42.6	0.6	1319	2	US-08-504-459-7	Sequence 24, App1
C 38	42.6	0.6	1493	1	US-08-340-820-24	Sequence 24, App1
C 39	42.6	0.6	1493	1	US-08-593-535-24	Sequence 24, App1
C 40	42.6	0.6	6124	4	US-08-213-419B-3	Sequence 3, App11
C 41	42.4	0.6	1882	4	US-09-370-253-1	Sequence 1, App11
C 42	42.2	0.6	2671	6	5168051-9	Patent No. 5168051
C 43	42	0.6	1114	4	US-09-152-060-41	Sequence 41, App1
C 44	42	0.6	1359	4	US-09-387-574-11	Sequence 11, App1
C 45	42	0.6	1359	4	US-09-668-096-11	Sequence 11, App1

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14
Query Match 1.0%; Score 62.4; DB 1; Length 7218;

OY	2320	CGTGTGCACCTCGAAGACATTTTGAAGAAAGTCATTACACAGAAACATTTGGTGAGCAT	2319
Db	4310	GTCTCCGACATTCAGCAATTTTATGCTAATGCTAAATGCTACTAATTTGTTCTC	4399
OY	2380	CTGAATTTCTCAGAGAAACAACACAGCTGAGAGAGATGAAGATTCATTAAAGCTTTTCA	2439
Db	4400	ATACCATTTGTTAAACACGCGAGATGTCAAGACACCTTCATCAATTTAACGCAATTTTGT	4459

RESULT 4
US-08-232-463-14


```

1 CLASSIFICATION: 435
2
3 ATTORNEY/AGENT INFORMATION:
4 NAME: BIFONT, GYLSSES J
5 REGISTRATION NUMBER: 39,908
6 REFERENCE/DOCKET NUMBER: DAM 398-9
7
8 TELECOMMUNICATION INFORMATION:
9
10 TELEPHONE: 410-671-1158
11
12 TELEFAX: 410-671-2534
13
14 INFORMATION FOR SEQ. ID NO: 5:
15
16 SEQUENCE CHARACTERISTICS:
17
18 LENGTH: 140 bases
19
20 TYPE: nucleic acid
21
22 STRANDEDNESS: single
23
24 TOPOLOGY: linear
25

```

MOLECULE TYPE: oligodeoxynucleotide
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-628-417-5

Query Match
Best Local Similarity 68.1%; Score 44.6; DB 1; Length 140;
Matches 62; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Y 5514 TTTTAAAGTATATTTCTGATTTTGTAAATTTAGAGTGTGTTGTTTGT 5573
D 113 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 54

Y 5574 TTTTGTATTTTATTTTAAAGAAACATTATTA 5604
D 53 TTTTATTTTATTTTATTTTAAATTTATTTCTAA 23

RESULT 13
US-08-628-417-6/C
Sequence 6, Application US/08628417
Patent No. 5627054
GENERAL INFORMATION:
APPLICANT: GILLESPIE, DAVID
TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
TITLE OF INVENTION: POLYMERASE CHAIN REACTION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
CITY: ABERDEEN PROVING GROUND
STATE: MARYLAND
COUNTRY: USA
ZIP: 21010-5423
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/628.417
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BIFFONT, ULXSES J
REGISTRATION NUMBER: 39,908
REFERENCE/DOCKET NUMBER: DAM 398-94
TELECOMMUNICATION INFORMATION:
TELEPHONE: 410-671-1158
TELEFAX: 410-671-2534
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: oligodeoxynucleotide
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-628-417-6

Query Match
Best Local Similarity 68.1%; Score 44.6; DB 1; Length 240;
Matches 62; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Y 5514 TTTTAAAGTATATTTCTGATTTTGTAAATTTAGAGTGTGTTTGT 5573
D 113 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 54

Y 5574 TTTTGTATTTTATTTTAAAGAAACATTATTA 5604
D 53 TTTTATTTTATTTTATTTTAAATTTATTTCTAA 23

RESULT 14
US-08-545-196B-10/C
Sequence 10, Application US/08545196B
Patent No. 6080577
GENERAL INFORMATION:
APPLICANT: MELKI, JUDITH
APPLICANT: MANNICH, ARNOLD
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545.196B
FILING DATE: 19-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FARACT, C. J.
REGISTRATION NUMBER: 32,350
REFERENCE/DOCKET NUMBER: 2121-110P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1582 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-545-196B-10

Query Match
Best Local Similarity 64.7%; Score 44.4; DB 3; Length 1582;
Matches 66; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Y 5514 TTTTAAAGTATATTTCTGATTTTGTAAATTTAGAGTGTGTTTGT 5573
D 1564 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1505

Y 5574 TTTTGTATTTTATTTTAAAGAAACATTATTAAGTGATGAT 5615
D 1504 TTTTATTTTATTTTATTTTAAATAAATTTATTTATTTATAT 1463

RESULT 15
US-08-545-196B-12/C
Sequence 12, Application US/08545196B
Patent No. 6080577
GENERAL INFORMATION:
APPLICANT: MELKI, JUDITH
APPLICANT: MANNICH, ARNOLD
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA

BLANK PAGE

XX WPI; 2000-499228/44.
DR P-PSDB; AAB07792.
DR

AA	Nucleic acids encoding guanine nucleotide releasing factor-4 useful for
PT	the treatment of cancers and neuronal disorders -

PS Claim 2; Fig 19A; 89pp; English.

The present sequence encodes a murine guanine nucleotide releasing factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both in vitro and in vivo. It directly binds cyclic adenosine monophosphate (cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cGMP) binding domain). GRF4 directly connects cAMP-generating (e.g. G protein coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4 is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of the Ras signalling pathway controls numerous cellular functions, such as cell metabolism, proliferation, differentiation and transformation. Therefore modulation of Ras activity may provide a mechanism for controlling diseases. GRF4 polynucleotides and polypeptides may be used in the treatment of diseases associated with inappropriate GRF4 expression and activity such as cancers and neuronal disorders. The GRF4 polypeptides may be used as antigens in the production of GRF4 antibodies against GRF4 and in assays to identify modulators (agonists and antagonists) of GRF4 expression and activity. The anti-GRF4 antibodies and GRF4 antagonists may also be used to down regulate GRF4 expression and activity. Inhibition of Ras can reduce cellulose proliferation and cancers.

Sequence 6568 BP; 1974 A; 1400 C; 1463 G; 1731 T; 0 other;

Query Match	100.0%;	Score 6568;	DB 21;	Length 6568;
Best Local Similarity	100.0%;	Pred. No. 0;		

Qy	1	CTTGCATCGTAGAGATTGGTACATGATGATGCTAAATCGAGTTCAGCATATGTTCTTCA	60
Db	1	CTTGCATCGTAGAGATTGGTACATGATGATGCTAAATCGAGTTCAGCATATGTTCTTCA	60
Qy	61	TTATGAAACCACTAGCAATCCGAGCTAAACCATGGAATTATGGCCAGCAGAGAAACACT	120
Db	61	TTATGAAACCACTAGCAATCCGAGCTAAACCATGGAATTATGGCCAGCAGAGAAACACT	120
Qy	121	CACCTCCTGCAGATTTCACAAATGCAATCTTACTGACAGTCTCCACCACAGGTGACC	180
Db	121	CACCTCCTGCAGATTTCACAAATGCAATCTTACTGACAGTCTCCACCACAGGTGACC	180
Qy	181	ACGTTTCTTCTTAGCCATTCCAGATGTAGTATCACTAGTGAATTCTGGAGCAGCAGTCTT	240
Db	181	ACGTTTCTTCTTAGCCATTCCAGATGTAGTATCACTAGTGAATTCTGGAGCAGCAGTCTT	240
Qy	241	CTGATTTCTACCAAGGCACAGAAACCGAGGCTGGTGATATGACCTGATGGGTGCGAG	300
Db	241	CTGATTTCTACCAAGGCACAGAAACCGAGGCTGGTGATATGACCTGATGGGTGCGAG	300
Qy	301	AAACACCACTGGATTCCGAGACGACGACGATGAGAGAAGACATTGAGAGACATCAGATC	360
Db	301	AAACACCACTGGATTCCGAGACGACGACGATGAGAGAAGACATTGAGAGACATCAGATC	360
Qy	361	CTGTGATGAGCAGGAGCATTTGTGAAGACGCGTATGAGAAAGACCCAAATTGACCGGACAG	420
Db	361	CTGTGATGAGCAGGAGCATTTGTGAAGACGCGTATGAGAAAGACCCAAATTGACCGGACAG	420
Qy	421	ATGATGACATTGAAACAATCTTGGAAATTTATGACACACTGCTGCTTTTGGCAATATGA	480
Db	421	ATGATGACATTGAAACAATCTTGGAAATTTATGACACACTGCTGCTTTTGGCAATATGA	480
Qy	481	CAATGTCAGTAGAGGAGAAACTCTGGCGTGTATGCTTGGCAGGTGGTGAAGAGACAG	540
Db	481	CAATGTCAGTAGAGGAGAAACTCTGGCGTGTATGCTTGGCAGGTGGTGAAGAGACAG	540
Qy	541	GGACCACTAGTGTAAATGATGGTGAAGACCTGACCTCGGTGCAATGATTCATAGAT	6000

Db	541	GGACCAATAGTGTAAATGATGGAGAGCTGGACTCTGGTCAGTATTTCTCAATGAT	600
Qy	601	CTGTGAAGAGACTTTATCCAGATGCAAAAGCAGAAATACCTGTGCATGGAAATAGTTTG	650
Db	601	CTGTGGAAAGTACTTATTCAGATGCAAAAGCAGAAATACCTGTGCATGGAAATAGTTTG	650
Qy	661	GTTGTCTCTCTACCATGAGACAAGATACATGAAGAGATGATGAGAACAAAGSTGATG	720
Db	661	GTTGTCTCTCTACCATGAGACAAGATACATGAAGAGATGATGAGAACAAAGSTGATG	720
Qy	721	ACTGCCAGTTTGTGCTGATGAGCCACAGATTTACTGCCGATTTCTCAATCAATAGAAA	780
Db	721	ACTGCCAGTTTGTGCTGATGAGCCACAGATTTACTGCCGATTTCTCAATCAATAGAAA	780
Qy	781	AGAACATGCAAAAAGTTGAAGAGAGAGAGATTTGTATGCTGAAGAACACCGAGAAC	840
Db	781	AGAACATGCAAAAAGTTGAAGAGAGAGAGATTTGTATGCTGAAGAACACCGAGAAC	840
Qy	841	TTGTATCCAACTGAGACAGAAAGAGGACACATTTGTCATCAAGGTCCTCACAAGGTTAA	900
Db	841	TTGTATCCAACTGAGACAGAAAGAGGACACATTTGTCATCAAGGTCCTCACAAGGTTAA	900
Qy	901	CAATGCAATTTGGTGGAGAGACATTTCAGTAGTAGATCCAAATTCATPAGAGACTTTCGT	960
Db	901	CAATGCAATTTGGTGGAGAGACATTTCAGTAGTAGATCCAAATTCATPAGAGACTTTCGT	960
Qy	961	TGACCTATAGACTTTTCTTCTTAGCCCAATGGAAGTGGGCAAAAAGTTATGGAGTGT	1020
Db	961	TGACCTATAGACTTTTCTTCTTAGCCCAATGGAAGTGGGCAAAAAGTTATGGAGTGT	1020
Qy	1021	TTTAATGACCCGAGCCTCAGGGATTAAGGTTACAGGGGTAGTATTATTTGGTAAATATC	1080
Db	1021	TTTAATGACCCGAGCCTCAGGGATTAAGGTTACAGGGGTAGTATTATTTGGTAAATATC	1080
Qy	1081	ACTTCATGACTTTTGAAAGAGATCTTGCAATGACTGCTGATTTTAAAGAAATTTGAAAAA	1140
Db	1081	ACTTCATGACTTTTGAAAGAGATCTTGCAATGACTGCTGATTTTAAAGAAATTTGAAAAA	1140
Qy	1141	ATCTGGAAGAGAGAAAATGGGTGGACACCTAAGGCTGTGAATATCGCGTGTGCTGTA	1200
Db	1141	ATCTGGAAGAGAGAAAATGGGTGGACACCTAAGGCTGTGAATATCGCGTGTGCTGTA	1200
Qy	1201	AAGCAAAAAGAGATGTGATCGTTAACAAACCAATCCCGAGAGACTCTTTGCCTTTGA	1260
Db	1201	AAGCAAAAAGAGATGTGATCGTTAACAAACCAATCCCGAGAGACTCTTTGCCTTTGA	1260
Qy	1261	TCTTACTTGGAGGCTGTGAGAAAGGATTTGGATCTTTGTGTACAGTGTAGATTCAGGTA	1320
Db	1261	TCTTACTTGGAGGCTGTGAGAAAGGATTTGGATCTTTGTGTACAGTGTAGATTCAGGTA	1320
Qy	1321	GCAAAAGCAACTGAAACGAGCTTGAACGGGGGAGTCAGATATTAGAATTAATGGCCAA	1380
Db	1321	GCAAAAGCAACTGAAACGAGCTTGAACGGGGGAGTCAGATATTAGAATTAATGGCCAA	1380
Qy	1381	ACCTTGAACAACTTAGCTGTCAAAACCTATGGAATTCCTTAAATTAACACACTTAT	1440
Db	1381	ACCTTGAACAACTTAGCTGTCAAAACCTATGGAATTCCTTAAATTAACACACTTAT	1440
Qy	1441	CTATCACTGTGAAAACCAATTTATTTGTATTTAAAGAACTTCAACAGATTTGTCAGAG	1500
Db	1441	CTATCACTGTGAAAACCAATTTATTTGTATTTAAAGAACTTCAACAGATTTGTCAGAG	1500
Qy	1501	AGAAAAGAAATGTGCCCCCACTCTCTAAATTTGGTGTGACATTTAAAGAGCCAGTCGT	1560
Db	1501	AGAAAAGAAATGTGCCCCCACTCTCTAAATTTGGTGTGACATTTAAAGAGCCAGTCGT	1560
Qy	1561	ACATCCATTCACATCTTGCTGTAGATGTAACAGATGATAGGACTTTGAAAAAGTGAACA	1620
Db	1561	ACATCCATTCACATCTTGCTGTAGATGTAACAGATGATAGGACTTTGAAAAAGTGAACA	1620
Qy	1621	AAAAAGTAAAGCCAACTGTGGAGAGAGAACAGCTGAAAAAGATATCTGCACACA	1680
Db	1621	AAAAAGTAAAGCCAACTGTGGAGAGAGAACAGCTGAAAAAGATATCTGCACACA	1680

Dd 1621 AAAAAAGTAAACCAACACTGTGGAGGAAAGAACACAGCTGAAAAAGATCTGCAACAGA 1680
Oy 1681 CTGGATCAGATATCTTGGCCACAGAAACCATACATGATATGGATGGTGCAGTCTCAAG 1740
Dd 1681 CTCGATCAGTATCTTGGCCACAGAAACCATACATGATATGGATGGTGCAGTCTCAAG 1740
Oy 1741 ATGACAGCATAGTAGATTAAGGACAGACAAAGACATCCCACTGCATGGCTGTCAAGT 1800
Dd 1741 ATGACAGCATAGTAGATTAAGGACAGACAAAGACATCCCACTGCATGGCTGTCAAGT 1800
Oy 1801 GAACCTTATCATTCCAGATATCTGATTTATTCAGTACATCATGCAATTTTATGACTTCA 1860
Dd 1801 GAACCTTATCATTCCAGATATCTGATTTATTCAGTACATCATGCAATTTTATGACTTCA 1860
Oy 1861 GTGCTACTCTCAGTTCGACATCAAGTCTTAAGGTTTATTAAGGCTGTATGCAAAAGCC 1920
Dd 1861 GTGCTACTCTCAGTTCGACATCAAGTCTTAAGGTTTATTAAGGCTGTATGCAAAAGCC 1920
Oy 1921 GCTACATCATGATCAGTAAAGACACTACAGCAAAAGAGTGTATTCAGGCTATCAGGG 1980
Dd 1921 GCTACATCATGATCAGTAAAGACACTACAGCAAAAGAGTGTATTCAGGCTATCAGGG 1980
Oy 1981 AGTTTGTCTTACTGCGCACCCCGGATCAATATTCTACTATGTAGGTTCTGTGCACACTG 2040
Dd 1981 AGTTTGTCTTACTGCGCACCCCGGATCAATATTCTACTATGTAGGTTCTGTGCACACTG 2040
Oy 2041 AGGAGTAACTCAACAAAGAGACATCCAGATCAGCTTCCAACTTGCAGACAGAAATAC 2100
Dd 2041 AGGAGTAACTCAACAAAGAGACATCCAGATCAGCTTCCAACTTGCAGACAGAAATAC 2100
Oy 2101 AACTGATGGAAGGATATTAATCTGAAAAACACATGAGAAAACAGAACTCTTTGTTCAGATG 2160
Dd 2101 AACTGATGGAAGGATATTAATCTGAAAAACACATGAGAAAACAGAACTCTTTGTTCAGATG 2160
Oy 2161 AAGATGCTCAGAGATGTTTGAGAGAGAGTCAATTTCCCTCTTCAAGCTCAGACACTGTGG 2220
Dd 2161 AAGATGCTCAGAGATGTTTGAGAGAGAGTCAATTTCCCTCTTCAAGCTCAGACACTGTGG 2220
Oy 2221 AAGTTGCAACACAGAGCTCTCTATGCGAAATTTTGAACCTTTTGCAGCAATTTGAACCTAGT 2280
Dd 2221 AAGTTGCAACACAGAGCTCTCTATGCGAAATTTTGAACCTTTTGCAGCAATTTGAACCTAGT 2280
Oy 2281 AATATATATAGATATTTTAACTCAGATCAAAAAACAGCTGTGGCAACCTGAAGAGAT 2340
Dd 2281 AATATATATAGATATTTTAACTCAGATCAAAAAACAGCTGTGGCAACCTGAAGAGAT 2340
Oy 2341 TTGAAGAAGTCTATTAACCGAGAAACATTTTGGGTAGCATCTGAAATTTCTCAGAGAAACAA 2400
Dd 2341 TTGAAGAAGTCTATTAACCGAGAAACATTTTGGGTAGCATCTGAAATTTCTCAGAGAAACAA 2400
Oy 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTATCATCAAGATAGACACTGAGG 2460
Dd 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTATCATCAAGATAGACACTGAGG 2460
Oy 2461 AATGCAAGAATTTTAACTCAATGTTTGAATCATCAGTGGCTTAAACCTGGACAGCAGTGG 2520
Dd 2461 AATGCAAGAATTTTAACTCAATGTTTGAATCATCAGTGGCTTAAACCTGGACAGCAGTGG 2520
Oy 2521 CAAGACTGGAGACAGCTGGGAGAACTTCCCAATTAATACAAAAAATCTATTTCAAGATC 2580
Dd 2521 CAAGACTGGAGACAGCTGGGAGAACTTCCCAATTAATACAAAAAATCTATTTCAAGATC 2580
Oy 2581 TCCAGAGACCTGTTTGAATCTTCCAGAAACATGCAAAATATGTAATGTTTCAATAGTC 2640
Dd 2581 TCCAGAGACCTGTTTGAATCTTCCAGAAACATGCAAAATATGTAATGTTTCAATAGTC 2640
Oy 2641 AAAATCTACACCTCCCATATTCCTTATTCAGATTATCAAAAAGATCTCACCTTCC 2700
Dd 2641 AAAATCTACACCTCCCATATTCCTTATTCAGATTATCAAAAAGATCTCACCTTCC 2700
Oy 2701 TTCACGAGAGAAATGACTCAAAAGTAGAGGGCTGTCAATTTTGAAGGCTAAGGATGA 2760
Dd 2701 TTCACGAGAGAAATGACTCAAAAGTAGAGGGCTGTCAATTTTGAAGGCTAAGGATGA 2760

Oy 2761 TTGCAAAAAGAAATGCTACGTTGGCCGATGCTTCAATGAACATGAGACCTGCCCCA 2820
Dd 2761 TTGCAAAAAGAAATGCTACGTTGGCCGATGCTTCAATGAACATGAGACCTGCCCCA 2820
Oy 2821 TGTTCAGACCTCGGAAGAAGAAATGGCGAGTTTGGGTCTCTCAGCCAGGGTAGTACAA 2880
Dd 2821 TGTTCAGACCTCGGAAGAAGAAATGGCGAGTTTGGGTCTCTCAGCCAGGGTAGTACAA 2880
Oy 2881 ATGCAACATGCTAGATGTTTGTCTCAGACAGGTGCTATAAAAAGCCGGTAGCTAGTT 2940
Dd 2881 ATGCAACATGCTAGATGTTTGTCTCAGACAGGTGCTATAAAAAGCCGGTAGCTAGTT 2940
Oy 2941 CCTTCTCATGCTCAAAAAAGCTTTATGAAGATGCCCAATGGCTCGAAAAGTGAACAGT 3000
Dd 2941 CCTTCTCATGCTCAAAAAAGCTTTATGAAGATGCCCAATGGCTCGAAAAGTGAACAGT 3000
Oy 3001 ACCTTTCCAAATTTGAGCTAGAAATGAGAGAGAGTCTTCAGACATTAATCTCTCAGT 3060
Dd 3001 ACCTTTCCAAATTTGAGCTAGAAATGAGAGAGAGTCTTCAGACATTAATCTCTCAGT 3060
Oy 3061 GTGAGCCAGCAACCAACACATTTGCTTAAGATCTGTGTACAAAAAGCCTGTCAATCCG 3120
Dd 3061 GTGAGCCAGCAACCAACACATTTGCTTAAGATCTGTGTACAAAAAGCCTGTCAATCCG 3120
Oy 3121 AGACCTCTCAGTAGCTCCAGAGGCGAGGTCACACAGAAAGCTCAGTCCGACAGC 3180
Dd 3121 AGACCTCTCAGTAGCTCCAGAGGCGAGGTCACACAGAAAGCTCAGTCCGACAGC 3180
Oy 3181 CCCAGCAGGACCCACACAGCAGCATMAATCAACCAAGGACTACAGTTCCCGCCGTG 3240
Dd 3181 CCCAGCAGGACCCACACAGCAGCATMAATCAACCAAGGACTACAGTTCCCGCCGTG 3240
Oy 3241 CCTTTATCTTCTCAGGAGAAAGTCCCGTAAAGATCTCCACCTTTTGGCATAACT 3300
Dd 3241 CCTTTATCTTCTCAGGAGAAAGTCCCGTAAAGATCTCCACCTTTTGGCATAACT 3300
Oy 3301 CTCACACACTTTAAAAAAATCTTTCTTGTCTGAAGAAGATTTTGAAGCTGCACA 3360
Dd 3301 CTCACACACTTTAAAAAAATCTTTCTTGTCTGAAGAAGATTTTGAAGCTGCACA 3360
Oy 3361 AGAAACAGGCTGAAGATACAAATATCAATGCAATCTTCGAGCTTCTCTCTACTT 3420
Dd 3361 AGAAACAGGCTGAAGATACAAATATCAATGCAATCTTCGAGCTTCTCTCTACTT 3420
Oy 3421 CTCACAGAGTCTCCAAAGAAAGCTATATCTTGGCTCCAGTGTACTGTGATTAAT 3480
Dd 3421 CTCACAGAGTCTCCAAAGAAAGCTATATCTTGGCTCCAGTGTACTGTGATTAAT 3480
Oy 3481 TTTTCAATCTGGTCCAGAGTAAATTTCTTACAGATCCAGTATTTGTAGCAATTCGCTT 3540
Dd 3481 TTTTCAATCTGGTCCAGAGTAAATTTCTTACAGATCCAGTATTTGTAGCAATTCGCTT 3540
Oy 3541 TTGACTCAGTGGCAGTCTCAGTCCAGATGAGAGGCGCCAGAGCAATTCGTGAGCATG 3600
Dd 3541 TTGACTCAGTGGCAGTCTCAGTCCAGATGAGAGGCGCCAGAGCAATTCGTGAGCATG 3600
Oy 3601 TGGAAACCAACCTAGGAGATGGGCAAGATGAGAGGCGGACATGATTTGAACCTGATCAGT 3660
Dd 3601 TGGAAACCAACCTAGGAGATGGGCAAGATGAGAGGCGGACATGATTTGAACCTGATCAGT 3660
Oy 3661 ATAGCTTGGGGTCTATGACCAATGTCGAGAGGCGGAGGCTATATGCTACAGCTACAG 3720
Dd 3661 ATAGCTTGGGGTCTATGACCAATGTCGAGAGGCGGAGGCTATATGCTACAGCTACAG 3720
Oy 3721 TAATTTCTTCTCCAAACAGAGAACTTCCAGAGATCAGGGGAGATCGCGCTCACTTG 3780
Dd 3721 TAATTTCTTCTCCAAACAGAGAACTTCCAGAGATCAGGGGAGATCGCGCTCACTTG 3780
Oy 3781 ATGCTGCTACAGATGGCGCTGGAGCTGAGAGCTCATGCTCAAGTGGCTCCATGATTAATA 3840
Dd 3781 ATGCTGCTACAGATGGCGCTGGAGCTGAGAGCTCATGCTCAAGTGGCTCCATGATTAATA 3840

QY	3841	TACAGACATCCAGCACCGAAGAGCTGGAGACTCTTCCATTCTGGGCATCTACTTTG	3900
Db	3841	TACAGAGCATCCAGACACCAGAGAGCTGGAGAGCTTCCATTCTGGGCATCTACTTTG	3900
QY	3901	ATTATTCAAGGGGATTCCTCAGGTTTATGGGGATCAAGCAGCATATGGACCAAAATTATG	3960
Db	3901	ATTATTCAAGGGGATTCCTCAGGTTTATGGCATCAGCAGCCATATGGACCAAAATTATG	3960
QY	3961	TTTTGATCATATAGCAACAAGTATTAACAGCAGAAATCAAAGTAGAGAGAGCCTTGAAACAG	4020
Db	3961	TTTTGATCATATAGCAACAAGTATTAACAGCAGAAATCAAAGTAGAGAGAGCCTTGAAACAG	4020
QY	4021	CCCAAGTCCCGAGCAAGCTCGGCGCTTCCACAGGTACTGGGGAGAAACATCAGAGGTG	4080
Db	4021	CCCAAGTCCCGAGCAAGCTCGGCGCTTCCACAGGTACTGGGGAGAAACATCAGAGGTG	4080
QY	4081	ACACAGGCACAAATAAAGCGAGGGGTGMAAGAGATGTTTCCATTGAAGCCGAACAGTA	4140
Db	4081	ACACAGGCACAAATAAAGCGAGGGGTGMAAGAGATGTTTCCATTGAAGCCGAACAGTA	4140
QY	4141	GGCTAAGCTCTGTGACTACGAAAGAAACCAACCTCTGCCCATTTGGCTGCCACATAGGTG	4200
Db	4141	GGCTAAGCTCTGTGACTACGAAAGAAACCAACCTCTGCCCATTTGGCTGCCACATAGGTG	4200
QY	4201	TGGCATCAAGTACTTACAAAGGGGCTATTGCACGAAGAGAGGGCAGGTATCGAAGCCCC	4260
Db	4201	TGGCATCAAGTACTTACAAAGGGGCTATTGCACGAAGAGAGGGCAGGTATCGAAGCCCC	4260
QY	4261	CGCCACACCCCTCCGGCTACATTGGAATTCCTACTACTTTCCAGAAGGGCACTGCC	4320
Db	4261	CGCCACACCCCTCCGGCTACATTGGAATTCCTACTACTTTCCAGAAGGGCACTGCC	4320
QY	4321	ATCCAGCCAGGAACCGCGCGATACAAAGCGTGGCCCTTGAGAGTCCGCGATGTGTGCAC	4380
Db	4321	ATCCAGCCAGGAACCGCGCGATACAAAGCGTGGCCCTTGAGAGTCCGCGATGTGTGCAC	4380
QY	4381	GATCCTTCGACAGAGCTGGGCTTTCATCGTACAGCAGCCACATGGGGCATCCACACAGA	4440
Db	4381	GATCCTTCGACAGAGCTGGGCTTTCATCGTACAGCAGCCACATGGGGCATCCACACAGA	4440
QY	4441	GCAGGGCTTGAAACAACTCTAGTGGCATTAACCGAAGCATGTGACCCGCGCTGGCCC	4500
Db	4441	GCAGGGCTTGAAACAACTCTAGTGGCATTAACCGAAGCATGTGACCCGCGCTGGCCC	4500
QY	4501	CTTATCAGTCCCAAGGGTTCCTCACCGAGAGAGTGAAGTGAACAGTTTCTGTGCTTT	4560
Db	4501	CTTATCAGTCCCAAGGGTTCCTCACCGAGAGAGTGAAGTGAACAGTTTCTGTGCTTT	4560
QY	4561	GAGGCACAGATTTTCTGGGAAGCAGAGCGAGCCACTGAAAGGAGACACAAAGAACGT	4620
Db	4561	GAGGCACAGATTTTCTGGGAAGCAGAGCGAGCCACTGAAAGGAGAGCACAAAGAACGT	4620
QY	4621	CCTGAGCATTTGGAGCCTTGGAACTCAGATTCGAGAGCGGTGACACAGTTTGCTCTTC	4680
Db	4621	CCTGAGCATTTGGAGCCTTGGAACTCAGATTCGAGAGCGGTGACACAGTTTGCTCTTC	4680
QY	4681	CCTGCGTTTAAAGCAGCATGGGGCTTCTCTCCCTTCCTTCCTTCCTTTGGCATGGA	4740
Db	4681	CCTGCGTTTAAAGCAGCATGGGGCTTCTCTCCCTTCCTTCCTTCCTTTGGCATGGA	4740
QY	4741	AATATGTGAAGAAATTCCTCTGGCAGCTTTCAGACTTTGTGTCTTGAATGACAGTGC	4800
Db	4741	AATATGTGAAGAAATTCCTCTGGCAGCTTTCAGACTTTGTGTCTTGAATGACAGTGC	4800
QY	4801	AGCAATCTTCCAGCTCCACTGTGTGCTGCCATCACACAGTATCATTTCCAAATTC	4860
Db	4801	AGCAATCTTCCAGCTCCACTGTGTGCTGCCATCACACAGTATCATTTCCAAATTC	4860
QY	4861	CAAGATCATCAACAAGATGATTAAGTCTGGGCTCAGCTTCTCAATGCCTGGAAGATTT	4920
Db	4861	CAAGATCATCAACAAGATGATTAAGTCTGGGCTCAGCTTCTCAATGCCTGGAAGATTT	4920
QY	4921	TTTTTAATCTTCCTTTTGAATTCATCAATCAAGTCTTGAAGCTTGAATCTATTGGGATAATG	4980

Db	4921	TTTTAAATCTCTCTTTTGATTTCAATCCAGTCTCCAGACTGTATCTCATTTGGATATATG	4980
Qy	4981	ACAAAAGCTAGCCATTGAACTACTCTGGGGCCCTTAAACCACCAAGAAAGACAAAGAAAAA	5040
Db	4981	ACAAAAGCTAGCCATTGAACTACTCTGGGGCCCTTAAACCACCAAGAAAGACAAAGAAAAA	5040
Qy	5041	CAATGAATCCTTGAGTACAGTCTGTCCACTGTGTACAATGCTCTCTTTAAAAA	5100
Db	5041	CAATGAATCCTTGAGTACAGTCTGTCCACTGTGTACAATGCTCTCTTTAAAAA	5100
Qy	5101	AAAAATGAGTTAAAGATTTTGTCAGAGATAATATATCCATTAAATGATTAAG	5160
Db	5101	AAAAATGAGTTAAAGATTTTGTCAGAGATAATATATCCATTAAATGATTAAG	5160
Qy	5161	TATATTTTAAACCTTAACTAGAGGTTGCCAGCTGTCTGAAAAACCAAAATATGCCGG	5220
Db	5161	TATATTTTAAACCTTAACTAGAGGTTGCCAGCTGTCTGAAAAACCAAAATATGCCGG	5220
Qy	5221	ACAGGGTGTGGCCACACCAAGAAAGAGGGAGACCGCTTGACCCCTGAGCTTCCATG	5280
Db	5221	ACAGGGTGTGGCCACACCAAGAAAGAGGGAGACCGCTTGAGCCCTGAGCTTCCATG	5280
Qy	5281	TCCCTGTGCTCTACCCCGGAAATGCGCTATCCTGGAGATGAATGTAGCCAAATTA	5340
Db	5281	TCCCTGTGCTCTACCCCGGAAATGCGCTATCCTGGAGATGAATGTAGCCAAATTA	5340
Qy	5341	TACCAAGACACCTCATCTGTCTCTCCAGAGATGGGGTCTTGTAAACGTGTTG	5400
Db	5341	TACCAAGACACCTCATCTGTCTCTCCAGAGATGGGGTCTTGTAAACGTGTTG	5400
Qy	5401	CACATGGCCAGGGAGAGAACTAGAACCTTGTCTCTGTAGCCCTTATGGAGCAGG	5460
Db	5401	CACATGGCCAGGGAGAGAACTAGAACCTTGTCTCTGTAGCCCTTATGGAGCAGG	5460
Qy	5461	ACGCTGTCATTTGGCGGATGTGTCTCTGCATGAGATGGATGGCAACCCATTTTAA	5520
Db	5461	ACGCTGTCATTTGGCGGATGTGTCTCTGCATGAGATGGATGGCAACCCATTTTAA	5520
Qy	5521	GTTATATTTCTTGATTTTGTTAATTTAGAGGTAGGTTTGTGTTTTGTTT	5580
Db	5521	GTTATATTTCTTGATTTTGTTAATTTAGAGGTAGGTTTGTGTTTTGTTT	5580
Qy	5581	TTTTTTTAAAGAAACATTTATACCTGGATAGCATTCAGTGAAGACAGCTTGGATGTT	5640
Db	5581	TTTTTTTAAAGAAACATTTATACCTGGATAGCATTCAGTGAAGACAGCTTGGATGTT	5640
Qy	5641	GGAGCTAATGCAAGCTGTTTACTGCTCTTTCAGACACCTCCCTTTATGAATTGGC	5700
Db	5641	GGAGCTAATGCAAGCTGTTTACTGCTCTTTCAGACAGCTCCCTTTATGAATTGGC	5700
Qy	5701	ATTAGGAATTAACAAGCTTTTAAACGTGATAAAAAGATCAAAAAACCTGGTTACACATGCC	5760
Db	5701	ATTAGGAATTAACAAGCTTTTAAACGTGATAAAAAGATCAAAAAACCTGGTTACATGCC	5760
Qy	5761	AGCCTTTCAGAGGAGGTTTATGTCACCAAGACTTAACCTCCAAAGTGGCTTATATGACGCTG	5820
Db	5761	AGCCTTTCAGAGGAGGTTTATGTCACCAAGACTTAACCTCCAAAGTGGCTTATATGACGCTG	5820
Qy	5821	CATATAGGAAGAGGCTTAAGTGTAGCAACCATCTGCTACAGCTGCTATTAACCTATAT	5880
Db	5821	CATATAGGAAGAGGCTTAAGTGTAGCAACCATCTGCTACAGCTGCTATTAACCTATAT	5880
Qy	5881	GACTGAATGACCCCTCCACTCTATTTTGTGTTGTTTGGACAGACTCCGGAAGAAAGTA	5940
Db	5881	GACTGAATGACCCCTCCACTCTATTTTGTGTTGTTTGGACAGACTCCGGAAGAAAGTA	5940
Qy	5941	AGAGTGCACCAATGTGATGTACTCAAAATGTAGAGAACTGCTGTGTTGGATTTTTTTTCC	6000
Db	5941	AGAGTGCACCAATGTGATGTACTCAAAATGTAGAGAACTGCTGTGTTGGATTTTTTTTCC	6000
Qy	6001	ATTAAATTCAGCTGATCATATTGATCAGTAGTAACGTAAATATGCTCAATTTTAA	6060
Db	6001	ATTAAATTCAGCTGATCATATTGATCAGTAGTAACGTAAATATGCTCAATTTTAA	6060

Db 6001 AATTAATTCAGCTGATCATATTGATCAGTAGATGAATGAACGTAAATAGCTCAAAATTTTAA 6060
Qy 6061 GCGGAATTCAGCTGTTTTCACCTGATCAACAAATGTCAGTCCTTATTATTATTTCT 6120
Db 6061 GCGGAATTCAGCTGTTTTCACCTGATCAACAAATGTCAGTCCTTATTATTATTTCT 6120
Qy 6121 CTCTGTCATCAGCAGCTTTGCTACTTCTGCTTATACATGTCATATGCAATTTGTAAT 6180
Db 6121 CTCTGTCATCAGCAGCTTTGCTACTTCTGCTTATACATGTCATATGCAATTTGTAAT 6180
Qy 6181 TTACATGTAATATGCAATTTGTCAGCTTTTATATATAGCTAGGACCTCATGTGCAT 6240
Db 6181 TTACATGTAATATGCAATTTGTCAGCTTTTATATATAGGCTATGACCTCATGTGCAT 6240
Qy 6241 ATAGAAAGCAGAAATCTGCTCTACCAAGTGCACAAATGTTCTGAAGCATTAAGT 6300
Db 6241 ATAGAAAGCAGAAATCTGCTCTACCAAGTGCACAAATGTTCTGAAGCATTAAGT 6300
Qy 6301 AATTGTAGAACATAGACATGCTAATCTCAGTTCGCTGCTGATGTCAGTGAATGTA 6360
Db 6301 AATTGTAGAACATAGACATGCTAATCTCAGTTCGCTGCTGATGTCAGTGAATGTA 6360
Qy 6361 CAATTACAGCTGATTTCTCTCATCTTGTGATCTACTTGTACCTGTATGCTTTAGAA 6420
Db 6361 CAATTACAGCTGATTTCTCTCATCTTGTGATCTACTTGTACCTGTATGCTTTAGAA 6420
Qy 6421 AGACATTTGGAGCTGATCCCTTTGTATTTTAAATCAATATTTGACATTTGGT 6480
Db 6421 AGACATTTGGAGCTGATCCCTTTGTATTTTAAATCAATATTTGACATTTGGT 6480
Qy 6481 TATATTTTGTGCAAGTGTAGAAAATGTAATGTAATGTTATGCTTTACATCCAGTTTGA 6540
Db 6481 TATATTTTGTGCAAGTGTAGAAAATGTAATGTAATGTTATGCTTTACATCCAGTTTGA 6540
Qy 6541 CAAGCTGGAATAATTAATATATACAT 6568
Db 6541 CAAGCTGGAATAATTAATATATACAT 6568

RESULT 2

ABN95037
ID ABN95037 standard; DNA; 6568 BP.

AC ABN95037;

DT 13-AUG-2002 (first entry)

DE Gene #1535 used to diagnose liver cancer.

KM Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;

KM metastatic liver tumour; cytostatic; expression profile; disease state;

KM disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

PN WO200229103-A2.

PD 11-APR-2002.

PF 02-OCT-2001; 2001WO-US30589.

PR 02-OCT-2000; 2000US-237054P.

PA (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI: 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer,

PT hepatocellular carcinoma or metastatic liver tumor in a patient,

PT involves detecting the level of expression of two or more genes in a

PT liver tissue sample

XX
PS Claim 1: SEQ ID NO 1535; 298bp; English.
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumor in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN95037-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytosolic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
SQ Sequence 6568 BP; 1974 A; 1400 C; 1463 G; 1731 T; 0 other:

Query Match 100.0%; Score 6568; DB 24; Length 6568;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGGCCATCGTGAGAGATTGTCATGATGTAATTCAGTTCAGACATATGTTCTTCA 60
Db 1 CTGGCCATCGTGAGAGATTGTCATGATGTAATTCAGTTCAGACATATGTTCTTCA 60
Qy 61 TTATGAACCCATAGCAATCCAGCTAACCATGAGATTATGGGCCACAGAGAAACACT 120
Db 61 TTATGAACCCATAGCAATCCAGCTAACCATGAGATTATGGGCCACAGAGAAACACT 120
Qy 121 CACTTCCTGCAATTTTCACAAATCTGATTTACTGACAGTCTCCACCCAGGTGACC 180
Db 121 CACTTCCTGCAATTTTCACAAATCTGATTTACTGACAGTCTCCACCCAGGTGACC 180
Qy 181 AGCTTCTCTACGCCAATTCGAGATGTACACTGATGTTCTGGAGAGACAGCTTT 240
Db 181 AGCTTCTCTACGCCAATTCGAGATGTACACTGATGTTCTGGAGAGACAGCTTT 240
Qy 241 CTGATATCTACAGGCGCACAGAAAGCGAGCTGTATATGACATGAGTGGTGGCAG 300
Db 241 CTGATATCTACAGGCGCACAGAAAGCGAGCTGTATATGACATGAGTGGTGGCAG 300
Qy 301 AAACAGCAGTGTGATTTCCGAAGACGAGCATGAGAAAGCATTTGAGAGACATCAGATC 360
Db 301 AAACAGCAGTGTGATTTCCGAAGACGAGCATGAGAAAGCATTTGAGAGACATCAGATC 360
Qy 361 CTCTGATGACGAGGACATTTGTAGAGAGCTGTAGAGAAAGGAGCCCAATGAGCGGACAG 420
Db 361 CTCTGATGACGAGGACATTTGTAGAGAGCTGTAGAGAAAGGAGCCCAATGAGCGGACAG 420
Qy 421 ATGATGACATTTGAACAACCTGGAATTTATGACACAGTTCCTGCTTTGCCAATATGA 480
Db 421 ATGATGACATTTGAACAACCTGGAATTTATGACACAGTTCCTGCTTTGCCAATATGA 480
Qy 481 CAATGTCAGTGAAGCGAAGAACTCTGTGCTGTATGCTGTTCGCAAGTGTGGAAGAGCAG 540
Db 481 CAATGTCAGTGAAGCGAAGAACTCTGTGCTGTATGCTGTTCGCAAGTGTGGAAGAGCAG 540
Qy 541 GGAACATAGTGTAAATGATGAGAGAGCTGACCTGCTGTCAGTATCTCAATGAT 600
Db 541 GGAACATAGTGTAAATGATGAGAGAGCTGACCTGCTGTCAGTATCTCAATGAT 600
Qy 601 CTGTGAAAGTGAATTCAGATGGAAGAAACAGAAATAGTGCATGGGAAATAGTTTG 660
Db 601 CTGTGAAAGTGAATTCAGATGGAAGAAACAGAAATAGTGCATGGGAAATAGTTTG 660
Qy 661 GTGTCTCTCTACCATGAGCAAAAGATATGAAAGGAGTATGAGAAAGGTGATG 720
Db 661 GTGTCTCTCTACCATGAGCAAAAGATATGAAAGGAGTATGAGAAAGGTGATG 720

OY	721	ACTGCGACATTTTGTCGTATAGCCAGCAAGATATCTGCGCTATTCTCAATCAAGTACAAA	780
Db	721	ACTGCGACATTTTGTCGTATAGCCAGCAAGATATCTGCGCTATTCTCAATCAAGTACAAA	780
OY	781	AGACATGCAAAAAGTTGAAGAGGAGAGATTGTTATGGTGAAGAACACCGAGAAC	840
Db	781	AGAACATGCAAAAAGTTGAAGAGGAGAGATTGTTATGGTGAAGAACACCGAGAAC	840
OY	841	TTGATCGAAACTGGAACAAGAAAGGACACATTTGTCATCAAGGGTACCTCAGAAAGGTTAA	900
Db	841	TTGATCGAAACTGGAACAAGAAAGGACACATTTGTCATCAAGGGTACCTCAGAAAGGTTAA	900
OY	901	CAATGACATTTGGTGGAAAGACATTAGTAGTAGTATCCAAATTCATAGAAAGCTTTCGT	960
Db	901	CAATGACATTTGGTGGAAAGACATTAGTAGTAGTATCCAAATTCATAGAAAGCTTTCGT	960
OY	961	TCACCATATAGACTTTTCTTCTTACGCCAATGGAAGTGGGCAAAAAGTTATTTGAGTGGT	1020
Db	961	TCACCATATAGACTTTTCTTCTTACGCCAATGGAAGTGGGCAAAAAGTTATTTGAGTGGT	1020
OY	1021	TTAATGACCCGAGCCTCAEGGATTAAGGTTACACGGGTAGTATTATTTGGGTGAATATAC	1080
Db	1021	TTAATGACCCGAGCCTCAEGGATTAAGGTTACACGGGTAGTATTATTTGGGTGAATATAC	1080
OY	1081	ACCTTCATGACTTTTGAAGAGATATCCGCAATGACTGATTTTGAAGAAATTTGAAAACA	1140
Db	1081	ACCTTCATGACTTTTGAAGAGATATCCGCAATGACTGATTTTGAAGAAATTTGAAAACA	1140
OY	1141	ATCTGGAAGAGAGAAATGGGTGACACCTTAAGGCTGTTGAATATCGGTGTGCTGCTA	1200
Db	1141	ATCTGGAAGAGAGAAATGGGTGACACCTTAAGGCTGTTGAATATCGGTGTGCTGCTA	1200
OY	1201	AACCAAAAAGAAATTTGATGACGTTTAACAAAACATTCGCCGAGAGCTCCTTGGCCTTTTA	1260
Db	1201	AACCAAAAAGAAATTTGATGATGACGTTTAACAAAACATTCGCCGAGAGCTCCTTGGCCTTTTA	1260
OY	1261	TCCTTACTTTGGAGGCTCTGAGAAAGGATTTTGGAAATCTTTGTTGACAGTGTAGATTCAAGTA	1320
Db	1261	TCCTTACTTTGGAGGCTCTGAGAAAGGATTTTGGAAATCTTTGTTGACAGTGTAGATTCAAGTA	1320
OY	1321	GCAAAAGCACTGAAAGCAGGCTTAAACGGGGGATCAGATATTAGAAGTAATAGGCCAAA	1380
Db	1321	GCAAAAGCACTGAAAGCAGGCTTAAACGGGGGATCAGATATTAGAAGTAATAGGCCAAA	1380
OY	1381	ACTTTGAAAACATTTACAGCTGTCAAAAAGCTATGGAATTTCTTGAATATACACACTTTTAT	1440
Db	1381	ACTTTGAAAACATTTACAGCTGTCAAAAAGCTATGGAATTTCTTGAATATACACACTTTTAT	1440
OY	1441	CTATCACTGTGAAACCAATTTATTTGTATTTAAGAACTTTACAAAGATTGTCCAGAG	1500
Db	1441	CTATCACTGTGAAACCAATTTATTTGTATTTAAGAACTTTACAAAGATTGTCCAGAG	1500
OY	1501	AGAAAGAAATGTGTGCCCCCACCCTTCTTAAATTTGGTGTACATTTAAAGGCCAATGCGCT	1560
Db	1501	AGAAAGAAATGTGTGCCCCCACCCTTCTTAAATTTGGTGTACATTTAAAGGCCAATGCGCT	1560
OY	1561	ACTGCAATTTCCAGATCTTGTGCTGTATGTATGACACAGGTGATATGACACTTGAAGAAAGTGACA	1620
Db	1561	ACTGCAATTTCCAGATCTTGTGCTGTATGTATGACACAGGTGATATGACACTTGAAGAAAGTGACA	1620
OY	1621	AAAAAAGTAAGCCCAACACTGTGTGGAGGAAGGAACAAGCTGAAAAAGATATCTGCACAGA	1680
Db	1621	AAAAAAGTAAGCCCAACACTGTGTGGAGGAAGGAACAAGCTGAAAAAGATATCTGCACAGA	1680
OY	1681	CTCGGATCAGTATCTTGGCCACAGAAACATTCATGATATTTGGATTTGGTCAAGTCAAG	1740
Db	1681	CTCGGATCAGTATCTTGGCCACAGAAACATTCATGATATTTGGATTTGGTCAAGTCAAG	1740
OY	1741	ATGACAGATGTAGATTAAAGGACAGCAAGACATATCCCAACTGCATTTGGCTGTCAAGT	1800
Db	1741	ATGACAGATGTAGATTAAAGGACAGCAAGACATATCCCAACTGCATTTGGCTGTCAAGT	1800
OY	1801	GAACTTTTCATCCAGTAATCCCTGATTTATTTGCAAGTCAATCATTCGCTTTTACACTTCA	1860

Db	1801	GAACCTTATCATCCAGTAATCCCTGATTTATTGACGTACATCATCCGATTTTAGACTTA	1860
Qy	1861	GTGCTACTCCTGACTTGGCCAGATCAAGTCGTAAAGGTTTTTAAAGCTGATCAGCAAAAGCC	1920
Db	1861	GTGCTACTCCTGACTTGGCCAGATCAAGTCGTAAAGGTTTTTAAAGGCTGATCAGCAAAAGCC	1920
Qy	1921	GCTACATCATGATCATGTAAGGACACTACAGCAAGGAAGTGTATTCAGGCTATCAGGG	1980
Db	1921	GCTACATCATGATCATGTAAGGACACTACAGCAAGGAAGTGTATTCAGGCTATCAGGG	1980
Qy	1981	AGTTGCTGTACTCGCCACCCCGGATCATTTTCACTATGTAGAGTCTCTGTACACCTG	2040
Db	1981	AGTTGCTGTACTCGCCACCCCGGATCATTTTCACTATGTAGAGTCTCTGTACACCTG	2040
Qy	2041	AGGGAGTAATCAAAACAAGAGACTTCCAGATCAGCTTTCCAAACCTTGACAGACAGATAC	2100
Db	2041	AGGGAGTAATCAAAACAAGAGACTTCCAGATCAGCTTTCCAAACCTTGACAGACAGATAC	2100
Qy	2101	AACTCAGTGGAAAGGTATTATCTGAAAAACACATGGAACGAAACCTTTGTTCAAGATG	2160
Db	2101	AACTCAGTGGAAAGGTATTATCTGAAAAACACATGGAACGAAACCTTTGTTCAAGATG	2160
Qy	2161	AAGATGCTCAGAGAGTTGTTGAGAGAGTCAAAATTTCCCTCCCTTACGCTACGACTGTGG	2220
Db	2161	AAGATGCTCAGAGAGTTGTTGAGAGAGTCAAAATTTCCCTCCCTTACGCTACGACTGTGG	2220
Qy	2221	AAGTTGCAACACAGCTCTCTATGCGGAAATTTTGAACCTCTTGCGCAACCTTAACCTACTG	2280
Db	2221	AAGTTGCAACACAGCTCTCTATGCGGAAATTTTGAACCTCTTGCGCAACCTTAACCTACTG	2280
Qy	2281	AATATATAGATGATTTTATTTAACTCAGATCAAAAAACCAGCTGTGCCAACCTGAAGAGAT	2340
Db	2281	AATATATAGATGATTTATTTAACTCAGATCAAAAAACCAGCTGTGCCAACCTGAAGAGAT	2340
Qy	2341	TTGAAAGACTATTACCAGSAAACATTTTGGGTAGCATCTGAAATTTCTGAGAAACAA	2400
Db	2341	TTGAAAGACTATTACCAGSAAACATTTTGGGTAGCATCTGAAATTTCTGAGAAACAA	2400
Qy	2401	ACCAGCTGAAGAGATGAAGATCATTAAGCATTTCAATCAAGATAGCAGCTGCTGAGGG	2460
Db	2401	ACCAGCTGAAGAGATGAAGATCATTAAGCATTTCAATCAAGATAGCAGCTGCTGAGGG	2460
Qy	2461	AATGCAAGATTTTAACTCAATGTTTGCAATCATCAGTGGCTTAAACCTGGACACAGTGG	2520
Db	2461	AATGCAAGATTTTAACTCAATGTTTGCAATCATCAGTGGCTTAAACCTGGACACAGTGG	2520
Qy	2521	CAGACACTGGGAAGCACTGGGAGAAACCTCCCAATAATATACGAAAAACATATTTCAAGATC	2580
Db	2521	CAGACACTGGGAAGCACTGGGAGAAACCTCCCAATAATATACGAAAAACATATTTCAAGATC	2580
Qy	2581	TCCAGAGCTGTTTATCTCTTCAGAAACATGCGAAAAATGCTAATGTTTCTCAATATGTC	2640
Db	2581	TCCAGAGCTGTTTATCTCTTCAGAAACATGCGAAAAATGCTAATGTTTCTCAATATGTC	2640
Qy	2641	AAATATACACCTGCCAATAATCCCTATATCCAGTATTAACAAAAGAGATCTCACCTTCC	2700
Db	2641	AAATATACACCTGCCAATAATCCCTATATCCAGTATTAACAAAAGAGATCTCACCTTCC	2700
Qy	2701	TTTCAGAAAGAAATGACTCAAAAGTAGACGGGCTGTCAAATTTTGGAAGCTPAAGGATGA	2760
Db	2701	TTTCAGAAAGAAATGACTCAAAAGTAGACGGGCTGTCAAATTTTGGAAGCTPAAGGATGA	2760
Qy	2761	TTTGCAAAAGAAATTTCTGTCACGTTTGCGCAATGAGCTTCAATGTAACATGACCTGCCCTCA	2820
Db	2761	TTTGCAAAAGAAATTTCTGTCACGTTTGCGCAATGAGCTTCAATGTAACATGACCTGCCCTCA	2820
Qy	2821	TGTTCAAGACTTCGGAAGAAGAAATGGCGAGTTTGAGGCTCTCTACCCAGAGGTAATACAA	2880
Db	2821	TGTTCAAGACTTCGGAAGAAGAAATGGCGAGTTTGAGGCTCTCTACCCAGAGGTAATACAA	2880
Qy	2881	ATGCAACAGTGTAGATGTTGCTCAGACAGGTGTCATAAAAACCGGTAACGTCGTAGTT	2940
Db	2881	ATGCAACAGTGTAGATGTTGCTCAGACAGGTGTCATAAAAACCGGTAACGTCGTAGTT	2940

Dh 2881 ATGCACAGTGTAGATGTTGCTCAGACAGTGTGTCATAAAAAGGGGTACGTGTAAGT 2940
Qy 2941 CCTTTCATATGCCAAAAGCTTTATGAAGATGCCCAAATGGCTCGAAAAGTGAAGAGT 3000
Dh 2941 CCTTTCATATGCCAAAAGCTTTATGAAGATGCCCAAATGGCTCGAAAAGTGAAGAGT 3000
Qy 3001 ACCTTTCAATTTGGAGCTGAATGAGCAGAGAGTCTTTCAGACATTTATCTCTGCAGT 3060
Dh 3001 ACCTTTCAATTTGGAGCTGAATGAGCAGAGAGTCTTTCAGACATTTATCTCTGCAGT 3060
Qy 3061 GTGAGCCAGCAGACACACATTTGCTTAAGAAATCTGTGTGACAAAAGCTGTCAATTCG 3120
Dh 3061 GTGAGCCAGCAGACACACATTTGCTTAAGAAATCTGTGTGACAAAAGCTGTCAATTCG 3120
Qy 3121 AGACCTTCACAGTGTCCAGAGGCGTCCAGACAGAAAAGCTCAATCCCTGTCACAGC 3180
Dh 3121 AGACCTTCACAGTGTCCAGAGGCGTCCAGACAGAAAAGCTCAATCCCTGTCACAGC 3180
Qy 3181 CCCAGCAGAGCCACACAGCAGACATTAATCAACAGAGGACTACAGGTTCCCGCTGT 3240
Dh 3181 CCCAGCAGAGCCACACAGCAGACATTAATCAACAGAGGACTACAGGTTCCCGCTGT 3240
Qy 3241 CCTTTATCTTCACAGGAGAAAGTCCCGTAAAGAGATCTCCACCTTTTGGCATAACT 3300
Dh 3241 CCTTTATCTTCACAGGAGAAAGTCCCGTAAAGAGATCTCCACCTTTTGGCATAACT 3300
Qy 3301 CTCACAGCTTTAAAAAATTTCTTTCTTTGTCTGAAGAAAGATTGGAAGTCA 3360
Dh 3301 CTCACAGCTTTAAAAAATTTCTTTCTTTGTCTGAAGAAAGATTGGAAGTCA 3360
Qy 3361 AGAAGAGGCTGAAGATACATATCAATGATGATCTTCGACCTTTCTCTCCCTACTT 3420
Dh 3361 AGAAGAGGCTGAAGATACATATCAATGATGATCTTCGACCTTTCTCTCCCTACTT 3420
Qy 3421 CTCACAGATTTCTCCAGAGAAAGGCTATACTTGGCTCCAGTGTGACTGTGATAAT 3480
Dh 3421 CTCACAGATTTCTCCAGAGAAAGGCTATACTTGGCTCCAGTGTGACTGTGATAAT 3480
Qy 3481 TTTAGATTTCTGGTCAAGTGAATTTCTTCAGATCAGATGTTAGCAATTCGCTT 3540
Dh 3481 TTTAGATTTCTGGTCAAGTGAATTTCTTCAGATCAGATGTTAGCAATTCGCTT 3540
Qy 3541 TTTAGATTTCTGGTCAAGTGAATTTCTTCAGATCAGATGTTAGCAATTCGCTT 3600
Dh 3541 TTTAGATTTCTGGTCAAGTGAATTTCTTCAGATCAGATGTTAGCAATTCGCTT 3600
Qy 3601 TGGAAACAAACCTAGGAGTGGGAGATGGAGGCGGACCATGATGAACCTGATCAGT 3660
Dh 3601 TGGAAACAAACCTAGGAGTGGGAGATGGAGGCGGACCATGATGAACCTGATCAGT 3660
Qy 3661 ATACTTGGGGTCTATGACATGATGTCGAGGGCGGCTTATATGCTACACTACAG 3720
Dh 3661 ATACTTGGGGTCTATGACATGATGTCGAGGGCGGCTTATATGCTACACTACAG 3720
Qy 3721 TAAATTTCTTCCAGAGCAGAGAACTTTCCAGAGATCAGGGGATGCGCGTCACTT 3780
Dh 3721 TAAATTTCTTCCAGAGCAGAGAACTTTCCAGAGATCAGGGGATGCGCGTCACTT 3780
Qy 3781 ATGCTGCTGACAGTGGCGGTGGAGCTGAGCTGATGCTCAAGTGGCTCCCATATA 3840
Dh 3781 ATGCTGCTGACAGTGGCGGTGGAGCTGAGCTGATGCTCAAGTGGCTCCCATATA 3840
Qy 3841 TACAGAGATTCACAGCAGAGAAAGCTGGGAGCTCTTCCATTTGGGATATCTCACTT 3900
Dh 3841 TACAGAGATTCACAGCAGAGAAAGCTGGGAGCTCTTCCATTTGGGATATCTCACTT 3900
Qy 3901 ATATATGAGGGGATCTGACAGTGTATGAGGATCAAGAGCAGATATGAGCAATTAAT 3960
Dh 3901 ATATATGAGGGGATCTGACAGTGTATGAGGATCAAGAGCAGATATGAGCAATTAAT 3960
Qy 3961 TTTCTGATCATAGCAAAAGTATTAACAGGCAAAATCAAAAGTAGAGAGAGCTTGAACAG 4020
Dh 3961 TTTCTGATCATAGCAAAAGTATTAACAGGCAAAATCAAAAGTAGAGAGAGCTTGAACAG 4020

Qy 4021 CCCAGTCCCGAGCAAGCTGGGCGTCTTCCACAGGTTACTGGGGAGAAAGACTCAGAGGTG 4080
Dh 4021 CCCAGTCCCGAGCAAGCTGGGCGTCTTCCACAGGTTACTGGGGAGAAAGACTCAGAGGTG 4080
Qy 4081 ACACAGCAGCAATTAAGAGGAGGAGTGAAGGATGTTTCCATTGAAGCCGAAGAGTA 4140
Dh 4081 ACACAGCAGCAATTAAGAGGAGGAGTGAAGGATGTTTCCATTGAAGCCGAAGAGTA 4140
Qy 4141 GCTTAAGTCTGTACTACGAGAAAGCAAGCTGTGCCATGTGCTGCCATATAGTG 4200
Dh 4141 GCTTAAGTCTGTACTACGAGAAAGCAAGCTGTGCCATGTGCTGCCATATAGTG 4200
Qy 4201 TGGCATCAAGTACTACAAAGGGGCTCAATGTCAGAAAGGAGGAGATGAGAGGCC 4260
Dh 4201 TGGCATCAAGTACTACAAAGGGGCTCAATGTCAGAAAGGAGGAGATGAGAGGCC 4260
Qy 4261 CGCCACCCCTCCCGGCTACATTTGGAATTTCCATTACTGACTTTCCAGAAAGGCACTCC 4320
Dh 4261 CGCCACCCCTCCCGGCTACATTTGGAATTTCCATTACTGACTTTCCAGAAAGGCACTCC 4320
Qy 4321 ATCCAGCCAGAAACCGCGGACTACAAAGTGGCCCTTCAGAGATGCGGATGTCGAC 4380
Dh 4321 ATCCAGCCAGAAACCGCGGACTACAAAGTGGCCCTTCAGAGATGCGGATGTCGAC 4380
Qy 4381 GATCCTCCGACACAGCTGGGCTTCATCCCTACAGCAGCCACATGGGATCCACACGA 4440
Dh 4381 GATCCTCCGACACAGCTGGGCTTCATCCCTACAGCAGCCACATGGGATCCACACGA 4440
Qy 4441 GCAGGCTGTGAACAAACCTAGTGGCATTAACAGCAAGCAATGACCCGCGCTCGCC 4500
Dh 4441 GCAGGCTGTGAACAAACCTAGTGGCATTAACAGCAAGCAATGACCCGCGCTCGCC 4500
Qy 4501 CTATATCACTCCCAAGGTTTTCACCGAGAGATGAAGATGAACAACTTCTGCTGTT 4560
Dh 4501 CTATATCACTCCCAAGGTTTTCACCGAGAGATGAAGATGAACAACTTCTGCTGTT 4560
Qy 4561 GAGGACAGACTTTTCTGGAAGCAGAGGAGGAGCCACTGAAGGAGAGCAAAAGAGT 4620
Dh 4561 GAGGACAGACTTTTCTGGAAGCAGAGGAGGAGCCACTGAAGGAGAGCAAAAGAGT 4620
Qy 4621 CCTGAGCATTTGAGGCTTGGAACTCAATTCAGAGAGGAGTGAACAGTATGCTCTTC 4680
Dh 4621 CCTGAGCATTTGAGGCTTGGAACTCAATTCAGAGAGGAGTGAACAGTATGCTCTTC 4680
Qy 4681 CCTGCTTTAAAGCAGCATGGGCTTCTTCCTCCCTTCCCTTGGCATGTGA 4740
Dh 4681 CCTGCTTTAAAGCAGCATGGGCTTCTTCCTCCCTTCCCTTGGCATGTGA 4740
Qy 4741 AATATGTGAAGAAATTCCTGGGCACTTTCAGACTTGTGCTTGAAGTGAACAGTGC 4800
Dh 4741 AATATGTGAAGAAATTCCTGGGCACTTTCAGACTTGTGCTTGAAGTGAACAGTGC 4800
Qy 4801 AGCAATCTTGGAGCTCCGACTGTGCTGCTGCGCATCAGACAGATATCATTTCCAAATTC 4860
Dh 4801 AGCAATCTTGGAGCTCCGACTGTGCTGCTGCGCATCAGACAGATATCATTTCCAAATTC 4860
Qy 4861 CAAGATCATCAACAAAGATATTCACCTGTGGTGCACATCTTCATATGCTGGAAAGATTT 4920
Dh 4861 CAAGATCATCAACAAAGATATTCACCTGTGGTGCACATCTTCATATGCTGGAAAGATTT 4920
Qy 4921 TTTTAAATCTTCTTTTGAATTTCAATCCAGTCTTCAAGCTTGAATGCGGTAATG 4980
Dh 4921 TTTTAAATCTTCTTTTGAATTTCAATCCAGTCTTCAAGCTTGAATGCGGTAATG 4980
Qy 4981 AGAAAGCTAGCACTTGAAGTCTTGGGCTTTAAACCCAGCAAGAGAGCAAAAGAAAA 5040
Dh 4981 AGAAAGCTAGCACTTGAAGTCTTGGGCTTTAAACCCAGCAAGAGAGCAAAAGAAAA 5040
Qy 5041 CAATGAATCTCTTGTAGTACAGTGTGCTTCCACTTGTTTTACAAATGCTCTTTTAAAA 5100
Dh 5041 CAATGAATCTCTTGTAGTACAGTGTGCTTCCACTTGTTTTACAAATGCTCTTTTAAAA 5100

QY 5101 AAAAAATGATTAAAGATTTTGTTCAGAGATTAATATATATCCATTATATGATACAG 5160
| | | | |
Db 5101 AAAAAATGATTAAAGATTTTGTTCAGAGATTAATATATATCCATTATATGATACAG 5160
QY 5161 TATTTATTTTAAACCTTAGTATGGGTTGCGCAGCTGGTTTCGAAAAACCAATATGCGG 5220
| | | | |
Db 5161 TATTTATTTTAAACCTTAGTATGGGTTGCGCAGCTGGTTTCGAAAAACCAATATGCGG 5220
QY 5221 ACAGGGTGTGGCCACACCAAGAAAGACGGGAACCTGGCTTGTGAACCTGGCTCCATG 5280
| | | | |
Db 5221 ACAGGGTGTGGCCACACCAAGAAAGACGGGAACCTGGCTTGTGAACCTGGCTCCATG 5280
QY 5281 TCCTTCGTGCTCACCCTGGAGAGTCCCTATTCCTGGAAGTATGAATGTAGCCATTAA 5340
| | | | |
Db 5281 TCCTTCGTGCTCACCCTGGAGAGTCCCTATTCCTGGAAGTATGAATGTAGCCATTAA 5340
QY 5341 TACCAAGACACCTCATCTGCTCTCCAGTGGATGGGTTCTCTGTAACCTGTTTG 5400
| | | | |
Db 5341 TACCAAGACACCTCATCTGCTCTCCAGTGGATGGGTTCTCTGTAACCTGTTTG 5400
QY 5401 CACATGGCCAGGAGGAGGAAGTACGACCTTGTCTGTCTGAGCCTTATGAGAGCAGG 5460
| | | | |
Db 5401 CACATGGCCAGGAGGAGGAAGTACGACCTTGTCTGTCTGAGCCTTATGAGAGCAGG 5460
QY 5461 ACAGGTGCATTTGGCGATGTGTCGCTGCATTTGAGATGAGATGGGAACCCCATTTTAA 5520
| | | | |
Db 5461 ACAGGTGCATTTGGCGATGTGTCGCTGCATTTGAGATGAGATGGGAACCCCATTTTAA 5520
QY 5521 GTTATATTTCTTGATTTTGTATTTAGAGGTAGGTTTGTGTTTGTGTTTGTGTT 5580
| | | | |
Db 5521 GTTATATTTCTTGATTTTGTATTTAGAGGTAGGTTTGTGTTTGTGTTTGTGTT 5580
QY 5581 TTTTATTAAGAAACATTTATTAAGTATGATGATGCGATGGAAGACCTTGGGATGTT 5640
| | | | |
Db 5581 TTTTATTAAGAAACATTTATTAAGTATGATGATGCGATGGAAGACCTTGGGATGTT 5640
QY 5641 GGAGCTAATGCGCAGCTGTTTATCTGCTTTCAAGACAGCCTCCCTTATTTGAATGGC 5700
| | | | |
Db 5641 GGAGCTAATGCGCAGCTGTTTATCTGCTTTCAAGACAGCCTCCCTTATTTGAATGGC 5700
QY 5701 ATTAGGAATTAACAAGCCTTTTAACGCTATTAAGATAAAAACCTGGTTAGACATGCG 5760
| | | | |
Db 5701 ATTAGGAATTAACAAGCCTTTTAACGCTATTAAGATAAAAACCTGGTTAGACATGCG 5760
QY 5761 AGCCTTGAAGGAGGAGTATGATCAGCAAGACTAACCTCAAGTGGCTTTATGAGAGCTG 5820
| | | | |
Db 5761 AGCCTTGAAGGAGGAGTATGATCAGCAAGACTAACCTCAAGTGGCTTTATGAGAGCTG 5820
QY 5821 CATATAGAGAAGGCTTAAGTATAGCAACCATCTGTCACAGCTGATTTAACCTTAAAT 5880
| | | | |
Db 5821 CATATAGAGAAGGCTTAAGTATAGCAACCATCTGTCACAGCTGATTTAACCTTAAAT 5880
QY 5881 GACTGAATATACCCCTTCACACTATTTTGTGTGTTTGGCAAGACTCCGGAAAAAGTGA 5940
| | | | |
Db 5881 GACTGAATATACCCCTTCACACTATTTTGTGTGTTTGGCAAGACTCCGGAAAAAGTGA 5940
QY 5941 AGGCTGCATTTAGTAGTACTCAATGTGAGAGAACTGCTGCTTGTGATTTTTTTTCC 6000
| | | | |
Db 5941 AGGCTGCATTTAGTAGTACTCAATGTGAGAGAACTGCTGCTTGTGATTTTTTTTCC 6000
QY 6001 ATTAATATTCAGCTGATCATATTCATCAGTATGAATTAACGTTCAAAATTTTAAAA 6060
| | | | |
Db 6001 ATTAATATTCAGCTGATCATATTCATCAGTATGAATTAACGTTCAAAATTTTAAAA 6060
QY 6061 GTGGAATTTGAGTGTTTTTCACGTATCAAAACATGTGCTTATATTAATTTCC 6120
| | | | |
Db 6061 GTGGAATTTGAGTGTTTTTCACGTATCAAAACATGTGCTTATATTAATTTCC 6120
QY 6121 CTTCCTGATCATGAGCATTTGTCTACTTGTCTTATTTACATTTGCAATTAATGATTTGAAT 6180
| | | | |
Db 6121 CTTCCTGATCATGAGCATTTGTCTACTTGTCTTATTTACATTTGCAATTAATGATTTGAAT 6180
QY 6181 TTACATGTAATATGATCATTAATTTGGCAGTTTATTTATATAGGCTATGAGCCTCATGTGCAT 6240
| | | | |

Db 6181 TTACATGTAATATGATCATTAATTTGGCAGTTTATTTATATAGGCTATGAGCCTCATGTGCAT 6240
| | | | |
QY 6241 ATAGAAAGACAGAAATCTACCTTACCACAGTTGCACAAATGTATTTACATTAAGT 6300
| | | | |
Db 6241 ATAGAAAGACAGAAATCTACCTTACCACAGTTGCACAAATGTATTTACATTAAGT 6300
QY 6301 AATGTGAACATAGAGACCTGATATCTAGTGGCTGTGATGTGACAGTGAATGTA 6360
| | | | |
Db 6301 AATGTGAACATAGAGACCTGATATCTAGTGGCTGTGATGTGACAGTGAATGTA 6360
QY 6361 CAATTAAGTGTGATTTCTCATACTTTTGTATCTACTTGTACCTGTATGCTTTTGAAG 6420
| | | | |
Db 6361 CAATTAAGTGTGATTTCTCATACTTTTGTATCTACTTGTACCTGTATGCTTTTGAAG 6420
QY 6421 AGACATTTGTGAGTCTGTATCCCTTTGTATTTTATATCATTAATTTGACATTTGGT 6480
| | | | |
Db 6421 AGACATTTGTGAGTCTGTATCCCTTTGTATTTTAAATCAATTAATTTGACATTTGGT 6480
QY 6481 TATATTTTGTGTAAGATGGTGAATGTACTATGTTATGCTTTCATCCAGTTTGTGA 6540
| | | | |
Db 6481 TATATTTTGTGTAAGATGGTGAATGTACTATGTTATGCTTTCATCCAGTTTGTGA 6540
QY 6541 CAAGCTGGAATAATTAATTAATTAACAT 6568
| | | | |
Db 6541 CAAGCTGGAATAATTAATTAATTAACAT 6568

RESULT 3
AAV84550
ID AAV84550 standard; DNA; 2890 BP.
XX
AC AAV84550;
XX
DE 01-MAR-1999 (first entry)
XX
DE Human secreted protein gene 140 clone HE2GT20.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
cognitive disorder; schizophrenia; prostatic; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
XX
PN W09854963-A2.
XX
PD 10-DEC-1998.
XX
PE 04-JUN-1998; 98WO-US11422.
XX
PR 18-DEC-1997; 97US-0070923.
PR 06-JUN-1997; 97US-0048877.
PR 06-JUN-1997; 97US-0048881.
PR 06-JUN-1997; 97US-0048884.
PR 06-JUN-1997; 97US-0048893.
PR 06-JUN-1997; 97US-0048896.
PR 06-JUN-1997; 97US-0048899.
PR 06-JUN-1997; 97US-0048915.
PR 06-JUN-1997; 97US-0048949.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048972.
PR 06-JUN-1997; 97US-0049020.
PR 06-JUN-1997; 97US-0049375.
PR 05-SEP-1997; 97US-0057628.
PR 05-SEP-1997; 97US-0057635.
PR 05-SEP-1997; 97US-0057644.
PR 05-SEP-1997; 97US-0057647.
PR 05-SEP-1997; 97US-0057650.

PR	05-SEP-1997	97US-0057661.
PR	05-SEP-1997	97US-0057667.
PR	05-SEP-1997	97US-0057761.
PR	05-SEP-1997	97US-0057764.
PR	05-SEP-1997	97US-0057770.
PR	05-SEP-1997	97US-0057775.
PR	05-SEP-1997	97US-0057778.
PR	06-JUN-1997	97US-0048875.
PR	06-JUN-1997	97US-0048878.
PR	06-JUN-1997	97US-0048882.
PR	06-JUN-1997	97US-0048885.
PR	06-JUN-1997	97US-0048894.
PR	06-JUN-1997	97US-0048897.
PR	06-JUN-1997	97US-0049000.
PR	06-JUN-1997	97US-0048916.
PR	06-JUN-1997	97US-0048962.
PR	06-JUN-1997	97US-0048970.
PR	06-JUN-1997	97US-0048974.
PR	06-JUN-1997	97US-0049373.
PR	05-SEP-1997	97US-0057584.
PR	05-SEP-1997	97US-0057629.
PR	05-SEP-1997	97US-0057642.
PR	05-SEP-1997	97US-0057645.
PR	05-SEP-1997	97US-0057648.
PR	05-SEP-1997	97US-0057651.
PR	05-SEP-1997	97US-0057652.
PR	05-SEP-1997	97US-0057668.
PR	05-SEP-1997	97US-0057762.
PR	05-SEP-1997	97US-0057765.
PR	05-SEP-1997	97US-0057771.
PR	05-SEP-1997	97US-0057776.
PR	06-JUN-1997	97US-0048876.
PR	06-JUN-1997	97US-0048880.
PR	06-JUN-1997	97US-0048883.
PR	06-JUN-1997	97US-0048892.
PR	06-JUN-1997	97US-0048895.
PR	06-JUN-1997	97US-0048898.
PR	06-JUN-1997	97US-0048901.
PR	06-JUN-1997	97US-0048917.
PR	06-JUN-1997	97US-0048963.
PR	06-JUN-1997	97US-0048971.
PR	06-JUN-1997	97US-0049015.
PR	06-JUN-1997	97US-0049374.
PR	05-SEP-1997	97US-0057627.
PR	05-SEP-1997	97US-0057634.
PR	05-SEP-1997	97US-0057643.
PR	05-SEP-1997	97US-0057646.
PR	05-SEP-1997	97US-0057649.
PR	05-SEP-1997	97US-0057654.
PR	05-SEP-1997	97US-0057666.
PR	05-SEP-1997	97US-0057760.
PR	05-SEP-1997	97US-0057763.
PR	05-SEP-1997	97US-0057769.
PR	05-SEP-1997	97US-0057777.
PR	05-SEP-1997	97US-0057777.

(HUMA-) HUMAN GENOME SCI INC.

Brewer LA, Carter KC, Dillon PJ, Edner R, Endress GA, Fan P, Feng P, Ferlie AM, Fitcher CL, Florence C, Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;

WPI; 1999-059865/05.
P-PSDB; AAW88673, AAW88983, AAW88984, AAW88985, AAW88986

New isolated human genes and the secreted polypeptides they encode are useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders

Claim 4; Page 400-402; 772pp; English.

The invention relates to nucleic acid sequences (AAV84411 to AAV84633) encoding human secreted proteins (AAW85334 to AAW8736). The secreted protein gene sequences are deposited with the ATCC under deposit numbers ATCC 91797, 91797A, 91797S, 91797T, 209077, 209078, 209079, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host cells comprising recombinant vectors containing the nucleic acid sequences are used for the recombinant production of the secreted proteins. The polynucleotide and amino acid sequences are useful for are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Pathological conditions can be also diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the polynucleotides, based on which tissues they are most highly expressed in, and include developing products for the diagnosis or treatment of cancer, neurodegenerative disorders, developmental abnormalities and foetal deficiencies, blood disorders, tumours, leukemias, diseases of the immune system, autoimmune diseases, hepatic and renal disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's and cognitive disorders, schizophrenia, restenosis, prostate diseases, obesity, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, diseases of testes, lung or thymus, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for identifying their binding partners. The present sequence represents a gene encoding a human secreted protein (see descriptor line for gene number and clone identification).

Sequence 2890 BP; 816 A; 631 C; 618 G; 821 T; 4 other:

Query Match	43.28;	Score 2840.2;	DB 20;	Length 2850;
Best Local Similarity	99.78;	Pred. No. 0;		
Matches 2862;	Conservative	3;	Mismatches 2;	Indels 3;
				Gaps 2;

3702 TTATATGCTACAGCTACAGTAATTTCTTCTCCAAGCACAGAGGAAGTTCCAGGATCAG 3761

DB 1 TTAATAGCTACAGCTACAGTAATTTCTTCTCCAGCACAGAGGANC TTCCAGGATCAG 60

3/82 GGGATCGCGCTCACTTGATGCTGCTGACAGTGGCCGTGGAGCTGGACGTCAATGCTCA 3821

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

—

[illegible][illegible]

3942 CATATGGACCAAAATTATGTTTCTGATCATAGCACAAAGTATAACAGGCTAAATACTAAAGT 4001

Db 241 CATATGACCAATTATGTTTCTGATCATACCAAAAGTATAACAGGCAAAATCAAGT 300

QY 4002 AGAGAGAGCCTTGAACACAGCCAGTCCCGAGCAAGCTGGGGCTCTTCCACAGGTACTGG 4061

db 301 AGAGAGAGCCTTGACACAAGCCAGTCCCGAGCAAGCTGGGGCTCTTCCACAGGTTACTGG 360

QY 4062 GGAGAGACTCAGAAGGTGACACAGGCACAATAAAGCGAGGGGTGGAAAGGATGTTCC 4121

Db 361 GGAGAAGACTCAGAAGGTGACACACAGGCCACAATAAAGCGGAGGGGTGGAAGGATGTTCC 420

41ZZ A1GAAAGCCGAAAGCAG1AGCC1AAGG1C1G1GAC1ACGGAAGAAACCAAGCC1G1CCCC 4181

[illegible]

[illegible]

Dh 541 GGCAGGTATTCGAGAGCCCGCCACCCCTCCCGGCTACATTGGAAATTCACCAATTAATCTGAC 600

QY 4302 TTTCAGAGGGCACTCCCATCCAGCCAGGAACCGCCGACTACAACTGGCCCTTCAG 4361

Db 601 TTTCCAGAAAGGCGACTCCCATCCAGCCAGAAACCCGCGAGCTACAACTGGCCCTTCAG 660
Qy 4362 AGATCGGGATGGTCCGACGATCCTCCGACACAGCTGGGCTTCATCCGACAGCA 4421
Db 661 AGATCGGGATGGTCCGACGATCCTCCGACACAGCTGGGCTTCATCCGACAGCA 720
Qy 4422 CATGGGATCCACACAGCAGCAGGCTGTGAAACCTCAGTGGCAATAACCGAAG 4481
Db 721 CATGGGATCCACACAGCAGGCTGTGAAACCTCAGTGGCAATAAAYGAAGAG 780
Qy 4482 TCTGACCCGCGCTCCGCTTCATCACTCCCAAGGTTTCCACGAGGAGTGAAGAT 4541
Db 781 TCTGACCCGCGCTCCGCTTCATCACTCCCAAGGTTTCCACGAGGAGTGAAGAT 840
Qy 4542 GAACAAAGTTCTGCTGTGAGGACAGACTTTCTGGAAGCAGAGGAGCACTGAAA 4601
Db 841 GAACAAAGTTCTGCTGTGAGGACAGACTTTCTGGAAGCAGAGGAGCACTGAAA 900
Qy 4602 GGAGAGCAACAAGAGAGCTCTGAGCATTTGAGGCTTGAACCTCACATCTGAGAGGT 4661
Db 901 GGAGAGCAACAAGAGAGCTCTGAGCATTTGAGGCTTGAACCTCACATCTGAGAGGT 960
Qy 4662 GGAGAGCTTCCCTCTCTCCCTGCTTAAAAAGCAGATGGGGCTTCTCTCCCTCTTC 4721
Db 961 GGAGAGCTTCCCTCTCTCCCTGCTTAAAAAGCAGATGGGGCTTCTCTCCCTCTTC 1020
Qy 4722 CTTCCTCCCTTGCATGTGAAATAGCTGTGAAGAAATTCCTGAGCTTTTCAGACTTGT 4781
Db 1021 CTTCCTCCCTTGCATGTGAAATAGCTGTGAAGAAATTCCTGAGCTTTTCAGACTTGT 1080
Qy 4782 TGCCTGAATGACAGCTGACATCTTCGAGCTCCCACTGCTGCTGCTCCGACATCAC 4841
Db 1081 TGCCTGAATGACAGCTGACATCTTCGAGCTCCCACTGCTGCTGCTCCGACATCAC 1140
Qy 4842 ACAGTATCTTCACAAATTCACAAGATCATCACAAAGATGATTACTCTGCTGCTGCTC 4901
Db 1141 ACAGTATCTTCACAAATTCACAAGATCATCACAAAGATGATTACTCTGCTGCTGCTC 1200
Qy 4902 TCAATGCTGGAAGATTTTTTTTATCTCTCTTTAGATTTTCAATCCAGTCCAGACT 4961
Db 1201 TCAATGCTGGAAGATTTTTTTTATCTCTCTTTAGATTTTCAATCCAGTCCAGACT 1260
Qy 4962 TCAATGCTGGAAGATTTTTTTTATCTCTCTTTAGATTTTCAATCCAGTCCAGACT 5021
Db 1261 TCAATGCTGGAAGATTTTTTTTATCTCTCTTTAGATTTTCAATCCAGTCCAGACT 1320
Qy 5022 CAAGGAACAAGAAACAAATGAATCTTTGAGTACAGTCTGCTGCTGCTGCTGCTGCT 5081
Db 1321 CAAGGAACAAGAAACAAATGAATCTTTGAGTACAGTCTGCTGCTGCTGCTGCTGCT 1380
Qy 5082 AATGCT 5139
Db 1381 AATGCT 1440
Qy 5140 AATGCT 5199
Db 1441 AATGCT 1500
Qy 5200 CTGAAACCAAAATATGCGGACAGGGGTGGCCACACCAAGAAACGGGAAGACTGGC 5259
Db 1501 CTGAAACCAAAATATGCGGACAGGGGTGGCCACACCAAGAAACGGGAAGACTGGC 1560
Qy 5260 TGTGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5319
Db 1561 TGTGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Qy 5320 TATGAATGTTAGCAATTAATACCAAGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5379
Db 1621 TATGAATGTTAGCAATTAATACCAAGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Qy 5380 GTTCTCTGTTAAAGCTGTTGACATGCGGAGGAGGAACTAGAACCTTGTGCTGCTGCT 5439
Db 5380 GTTCTCTGTTAAAGCTGTTGACATGCGGAGGAGGAACTAGAACCTTGTGCTGCTGCT 5439

Db 1681 GTTCTCTGTTAAAGCTGTTGACATGCGGAGGAGGAACTAGAACCTTGTGCTGCTGCT 1740
Qy 5440 TGTGACCTTATGAGGACGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5499
Db 1741 TGTGACCTTATGAGGACGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Qy 5500 GATGCAAAACCCATTTTATGATTAATTTGATTTTGTATTTTGTATTTTGTATTTTGTATTT 5559
Db 1801 GATGCAAAACCCATTTTATGATTAATTTGATTTTGTATTTTGTATTTTGTATTTTGTATTT 1860
Qy 5560 TTTTGTGTTTGTG-TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 5618
Db 1861 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 1920
Qy 5619 AGTGAAGAGGCTGGGATGTTGAGCTAATGCGAGCTGTTTATAGCTGCTGCTTCAAGAC 5678
Db 1921 AGTGAAGAGGCTGGGATGTTGAGCTAATGCGAGCTGTTTATAGCTGCTGCTTCAAGAC 1980
Qy 5679 AGCTTCCCTTTTATGAAATGCGATTAGGAAATTAACAAACCTTTTAACTGATTAAGAT 5738
Db 1981 AGCTTCCCTTTTATGAAATGCGATTAGGAAATTAACAAACCTTTTAACTGATTAAGAT 2040
Qy 5739 CAATAACCTGTTAGACATGCGAGCTTTGCAAGGCGAGTTAGTACCAAGACTAACCT 5798
Db 2041 CAATAACCTGTTAGACATGCGAGCTTTGCAAGGCGAGTTAGTACCAAGACTAACCT 2100
Qy 5799 CCAAGTGGCTTATGAGGCTGCTGATAGAGGCTTAAGTGAAGCAACATCTGCTCA 5858
Db 2101 CCAAGTGGCTTATGAGGCTGCTGATAGAGGCTTAAGTGAAGCAACATCTGCTCA 2160
Qy 5859 CAGCTGCTATTAACCTTATGATGACTGAATGACCCCTGCTCATTTTGTGTTGTT 5918
Db 2161 CAGCTGCTATTAACCTTATGATGACTGAATGACCCCTGCTCATTTTGTGTTGTT 2220
Qy 5919 TGCACAGACTCCGGAAGAGGAGGCTGCAATCTGAGTACACAAATGAGAGAACT 5978
Db 2221 TGCACAGACTCCGGAAGAGGAGGCTGCAATCTGAGTACACAAATGAGAGAACT 2280
Qy 5979 GCTGCTGCTGATTTTTTTTCCATTAATTCAGCTGATCATTTGATGATGATTAAGC 6038
Db 2281 GCTGCTGCTGATTTTTTTTCCATTAATTCAGCTGATCATTTGATGATGATTAAGC 2340
Qy 6039 TAAATACCTTCAAAATTTTAAAGTGAATGAGTGTGTTTTCACGCTATCAACAAGT 6098
Db 2341 TAAATACCTTCAAAATTTTAAAGTGAATGAGTGTGTTTTCACGCTATCAACAAGT 2400
Qy 6099 CAGTGTGTTTATTAATTAATTCCTCTGTAATGAGCAATTTGCTAGTCTTATTAAT 6158
Db 2401 CAGTGTGTTTATTAATTAATTCCTCTGTAATGAGCAATTTGCTAGTCTTATTAAT 2460
Qy 6159 TGTCAATTAATGATTTTGAATTTTACATGTAATGATTAATGCTGCTGCTGCTGCTGCT 6218
Db 2461 TGTCAATTAATGATTTTGAATTTTACATGTAATGATTAATGCTGCTGCTGCTGCTGCT 2520
Qy 6219 AGGCTATGAGGCTGATGATATGAAAGACAGAAATGAGCTTACCAACAAGTGGAC 6278
Db 2521 AGGCTATGAGGCTGATGATATGAAAGACAGAAATGAGCTTACCAACAAGTGGAC 2580
Qy 6279 AAATGTTATCAAGCAATTAAGTAATGTAACATAGAGACTGCTAATCTCAGTCTGCT 6338
Db 2581 AAATGTTATCAAGCAATTAAGTAATGTAACATAGAGACTGCTAATCTCAGTCTGCT 2640
Qy 6339 GTGATGTCAGTGCAGATGTCAGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 6398
Db 2641 GTGATGTCAGTGCAGATGTCAGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
Qy 6399 TGTACCTGATGCTTTTGAAGACACTGCTGAGTCTGATCCCTTTTGTATTTTAA 6458
Db 2701 TGTACCTGATGCTTTTGAAGACACTGCTGAGTCTGATCCCTTTTGTATTTTAA 2760
Qy 6459 TACATAATTTGATACATATTTGTTATTTTGTGAAAGATGTAAGAAATGATGATGTT 6518
Db 2761 TACATAATTTGATACATATTTGTTATTTTGTGAAAGATGTAAGAAATGATGATGTT 2820

OY 6519 ATGCTTACATCCAGTTGTACAGCTGGAAAATAAATAATATAACAT 6568
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 2821 ATGCTTACATCCAGTTGTACAGCTGGAAAATAAATAATATAACAT 2870

RESULT 4
 ABA83333
 ID ABA83333 standard; cDNA; 2890 BP.
 XX
 AC ABA83333;
 XX
 DT 07-FEB-2002 (first entry)
 XX

Human secreted protein gene 140 SEQ ID NO:150.

Human: secreted protein; immunomodulatory; antisclerotic; anti-HIV;
 dermatological; immunosuppressive; antiinflammatory; immunostimulant;
 cytosolic; cardiant; vascular; anti-angiogenic; ophthalmological;
 neuroprotective; neurotropic; anticonvulsant; antialzheimers; vulnerary;
 antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;
 multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;
 human immunodeficiency virus; hyperproliferative disorder; wound healing;
 Gaucher's disease; cardiovascular disease; Scimitar syndrome; chemotaxis;
 Chaga's cardiomyopathy; coronary arteriosclerosis; angiogenic disorder;
 corneal graft neovascularisation; diabetic retinopathy; regeneration;
 neurological disorder; Huntington's chorea; Alzheimer's disease;
 Parkinson's disease; infectious disease; chromosome 4; ss.

Homo sapiens.

MO200162891-A2.

30-AUG-2001.

21-FEB-2001; 2001WO-US05614.

24-FEB-2000; 2000US-184836P.

29-MAR-2000; 2000US-193170P.

(HUMA-) HUMAN GENOME SCI INC.

NI J, Ebner R, Lafleur DW, Moore PA, Olsen HS, Rosen CA;

Ruben SM, Soppet DR, Young PE, Shi Y, Florence KA, Wei Y;

Florence C, Hu J, Li Y, Kyaw H, Fischer CL, Ferrle AM, Fan P;

Feng P, Endress GA, Dillon PJ, Carter KC, Brewer LA, Yu G;

Zeng Z, Greene JM;

WPI: 2001-625724/72.

P-PSDB; ABB50440.

Claim 1: Page 1004-1005; 1533pp; English.

ABBS0301 to ABB51287 and ABA83194 to ABA83441 represent human secreted

proteins (I) and polynucleotide (II) sequences. (I) and (II) have various

activities based on the tissues and cells the genes are expressed in.

Example of these activities include: immunomodulatory; antisclerotic;

dermatological; immunosuppressive; antiinflammatory; immunostimulant;

anti-HIV; cytosolic; cardiant; anti-angiogenic; ophthalmological;

neuroprotective; neurotropic; anticonvulsant; antialzheimers; vascular;

antiparkinsonian; antimicrobial; and vulnerary. (I) and (II) can be used

in gene therapy and vaccine production. (I) and (II) can be used in the

prevention, diagnosis and treatment of immune disorders (e.g. multiple

CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or

CC for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to

CC ABA83193 and ABB50300 represent sequences used in the exemplification of

CC the present invention.

XX

Sequence 2890 BP; 816 A; 631 C; 618 G; 821 T; 4 other;

Query Match 43.2%; Score 2840.2; DB 22; Length 2890;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 2862; Conservative 3; Mismatches 2; Indels 3; Gaps 2;

OY 3702 TTATATGCTACAGCTACAGTAATTTCTTCACAGCAGACAGAACTTCCAGATCAG 3761

||||| 1 TTATATGCTACAGCTACAGTAATTTCTTCACAGCAGACAGAACTTCCAGATCAG 60

OY 3762 GGGGATCGCGCTGCTACCTGATGCTGCTGACAGTGGCCGTGGAGCTGCATGCTCA 3821

||||| 61 GGGGATCGCGCTGCTACCTGATGCTGCTGACAGTGGCCGTGGAGCTGCATGCTCA 120

OY 3822 AGTGGCTCCCATATATATATACAGAGATCCACACAGAAAGCTGGAGACTCTTCCA 3881

||||| 121 AGTGGCTCCCATATATATATACAGAGATCCACACAGAAAGCTGGAGACTCTTCCA 180

OY 3882 TTGGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3941

||||| 181 TTGGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240

OY 3942 CATATGAGCAAAATTTATGTTTCTGATCATAGCACAAGTAATTAACAGGCAAAATCAAGT 4001

||||| 241 CATATGAGCAAAATTTATGTTTCTGATCATAGCACAAGTAATTAACAGGCAAAATCAAGT 300

OY 4002 AGAGAGAGCTTTAACAAGCCAGTCCGAGCAAGCTGGGCTCTTCCACAGTTACTGG 4061

||||| 301 AGAGAGAGCTTTAACAAGCCAGTCCGAGCAAGCTGGGCTCTTCCACAGTTACTGG 360

OY 4062 GGAGAAAGCTCAGAAAGTGCACAGGCAATTAAGGAGGAGGAGGAGGAGGAGGAGGAGG 4121

||||| 361 GGAGAAAGCTCAGAAAGTGCACAGGCAATTAAGGAGGAGGAGGAGGAGGAGGAGGAGG 420

OY 4122 ATTGAAGCCGAAAGCAGTACCTTAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4181

||||| 421 ATTGAAGCCGAAAGCAGTACCTTAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

OY 4182 ATGCTGCTCCACATAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4241

||||| 481 ATGCTGCTCCACATAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

OY 4242 GGCAGGATGAGAGGAGCCCGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4301

||||| 541 GGCAGGATGAGAGGAGCCCGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600

OY 4302 TTTTCAGAAAGGAGCTCCATCCAGCAGCAGGAAACCGCGGAGCTACAAAGTGGCCCTTCAG 4361

||||| 601 TTTTCAGAAAGGAGCTCCATCCAGCAGCAGGAAACCGCGGAGCTACAAAGTGGCCCTTCAG 660

OY 4362 AGATCGGAGATGCTGCGACATGCTGCGACAGAGCTGGGCTTCATCGTACAGACCCA 4421

||||| 661 AGATCGGAGATGCTGCGACATGCTGCGACAGAGCTGGGCTTCATCGTACAGACCCA 720

OY 4422 CATGGGATCCGACAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4481

||||| 721 CATGGGATCCGACAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780

OY 4482 TCTGACCCGCGCTGCGCTTATCATGCTCCAGAGGAGTTCCTCCAGAGAGGAGTGAAGAT 4541

||||| 781 TCTGACCCGCGCTGCGCTTATCATGCTCCAGAGGAGTTCCTCCAGAGAGGAGTGAAGAT 840

OY 4542 GAACAGTTTCTGCTGTTTGAAGCAGAGACTTTTGGAAAGCAGAGGAGGAGGAGGAGGAG 4601

||||| 841 GAACAAATTTCTGCTGTTTGAAGCAGAGACTTTTGGAAAGCAGAGGAGGAGGAGGAGGAG 900

||||| 4602 GGAGAGCACAAGAGAGCTGCTGAGCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4661

Db 901 GGAGAGCACAAGAGAGCTCTGAGCATGGAGCCTTGGAACTCACAATTCTGACGACGCT 960
Qy 4662 GGAGCAGTTTGGCCCTCTTCCCTGCTTAAAGCAGAGAGGGGCTCTCTCCCTTCTTC 4721
Db 961 GGAGCAGATTGGCTCTCTCTCCCTGCTTAAAGCAGAGATGGGGSTCTCTCTCCCTCTTC 1020
Qy 4722 CTTTCCCTTGGCATGTGAATAGTGTGAAGAAATGGCCCTGGCAGCTTTTCAGACTTTGT 4781
Db 1021 CTTTCCCTTGGCATGTGAATAGTGTGAAGAAATGGCCCTGGCAGCTTTTCAGACTTTGT 1080
Qy 4782 TGCCTTGAATAGCAGAGTGCAGCATCTTGGAGCTCCCACTGTGGCTGCTGCACATCAC 4841
Db 1081 TGCCTTGAATAGCAGAGTGCAGCATCTTGGAGCTCCCACTGTGGCTGCTGCACATCAC 1140
Qy 4842 ACAGTATCATTCCAAAATTCACAGATCATCACACAGATGATTCAGTGTGGCTGCTGCAC 4901
Db 1141 ACAGTATCATTCCAAAATTCACAGATCATCACACAGATGATTCAGTGTGGCTGCTGCAC 1200
Qy 4902 TCATGCTGGAAGATTTTCTTAACTCTCTTGTAGATTCAATCCAGCTGACACT 4961
Db 1201 TCATGCTGGAAGATTTTCTTAACTCTCTTGTAGATTCAATCCAGCTGACACT 1260
Qy 4962 TCATGCTGGAAGATTTTCTTAACTCTCTTGTAGATTCAATCCAGCTGACACT 5021
Db 1261 TCATGCTGGAAGATTTTCTTAACTCTCTTGTAGATTCAATCCAGCTGACACT 1320
Qy 5022 CAAGGAGACAAAGAAACAGTAACTCTTGGATGACAGTGTGGCTGCTGCACCTTTTAC 5081
Db 1321 CAAGGAGACAAAGAAACAGTAACTCTTGGATGACAGTGTGGCTGCTGCACCTTTTAC 1380
Qy 5082 AATGCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 5139
Db 1381 AATGCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1440
Qy 5140 AATGCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 5199
Db 1441 AATGCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1500
Qy 5200 CTGAAACCAAAATATGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5259
Db 1501 CTGAAACCAAAATATGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
Qy 5260 TTGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5319
Db 1561 TTGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
Qy 5320 TATGAATGTATGACCAATTAATACCAAGACACCTCATCTGCTGCTGCTGCTGCTGCTG 5379
Db 1621 TATGAATGTATGACCAATTAATACCAAGACACCTCATCTGCTGCTGCTGCTGCTGCTG 1680
Qy 5380 GTTCTTCTGTAATACTGTTGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5439
Db 1681 GTTCTTCTGTAATACTGTTGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
Qy 5440 TCTGAGCCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5499
Db 1741 TCTGAGCCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
Qy 5500 GATGCGAAACCCCATTTTAATTAATTTCTGATTTTGTATTAATTAAGAGTGTAGG 5559
Db 1801 GATGCGAAACCCCATTTTAATTAATTTCTGATTTTGTATTAATTAAGAGTGTAGG 1860
Qy 5560 TTTTGTGTTTGTG-TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 5618
Db 1861 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1920
Qy 5619 AGTGAAGAGAGCTTGGAGTGTGGAGCTAATGCCAGCTGTTTACTGCTCTTTCAGAGC 5678
Db 1921 AGTGAAGAGAGCTTGGAGTGTGGAGCTAATGCCAGCTGTTTACTGCTCTTTCAGAGC 1980
Qy 5679 AGCCTCCCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5738
Db 1981 AGCCTCCCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2040

Qy 5739 CAANAACCTGGTTAGACATGCCAGCCTTTGCAAGGACAGTTAGTACACAAAGACTAAGCT 5798
Db 2041 CAANAACCTGGTTAGACATGCCAGCCTTTGCAAGGACAGTTAGTACACAAAGACTAAGCT 2100
Qy 5799 CCAGTGGCTTATGAGGAGCTGCATATAGAGAGGCTTAAGTGTAGACACATCTGCTCA 5858
Db 2101 CCAGTGGCTTATGAGGAGCTGCATATAGAGAGGCTTAAGTGTAGACACATCTGCTCA 2160
Qy 5859 CAGCTGCTATTAACCTTAATAGATGAATAGACCCCTGCACTATTTTGTGTTGTTT 5918
Db 2161 CAGCTGCTATTAACCTTAATAGATGAATAGACCCCTGCACTATTTTGTGTTGTTT 2220
Qy 5919 TGCACAGACTCCGGAAGAGTGAAGCTGCCAATCTGAGTACTCAAAATGTAGAGAACT 5978
Db 2221 TGCACAGACTCCGGAAGAGTGAAGCTGCCAATCTGAGTACTCAAAATGTAGAGAACT 2280
Qy 5979 GCTGCTGCTGAGATTTTCTTCAATTAATTCAGCTGATCATATTTATGATAGATTAAG 2340
Db 2281 GCTGCTGCTGAGATTTTCTTCAATTAATTCAGCTGATCATATTTATGATAGATTAAG 2400
Qy 6039 TAATAGCTTCAAAATTTTAAAGTGAATTCAGTGTGTTTCTGCTGATCAAAACATGT 6098
Db 2341 TAATAGCTTCAAAATTTTAAAGTGAATTCAGTGTGTTTCTGCTGATCAAAACATGT 2400
Qy 6099 CAGTGTATTAATTAATTAATTTCTGCTGATGATGAGCTTGTCTGATGCTTATTAAT 6158
Db 2401 CAGTGTATTAATTAATTAATTTCTGCTGATGATGAGCTTGTCTGATGCTTATTAAT 2460
Qy 6159 TGTCAATTAATGATTTGTAATTTTACATGATTAATGATTAATTTGCAAGTTTATTAAT 6218
Db 2461 TGTCAATTAATGATTTGTAATTTTACATGATTAATGATTAATTTGCAAGTTTATTAAT 2520
Qy 6219 AGGCTATGACACATGATGATATAGAAAGACAGAAATCTAGCTCTACACAAAGTTGCAC 6278
Db 2521 AGGCTATGACACATGATGATATAGAAAGACAGAAATCTAGCTCTACACAAAGTTGCAC 2580
Qy 6279 AATGCTATTAATGATTAATTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAAT 6338
Db 2581 AATGCTATTAATGATTAATTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAAT 2640
Qy 6339 GTGATGCAAGTGAAGATGATTAATTAATTTGTAATTTGTAATTTGTAATTTGTAATTTG 6398
Db 2641 GTGATGCAAGTGAAGATGATTAATTAATTTGTAATTTGTAATTTGTAATTTGTAATTTG 2700
Qy 6399 TGTACCTGATGCTTTTGAAGACATTTGGTGAAGTGTATATCCCTTTGTAATTTTGA 6458
Db 2701 TGTACCTGATGCTTTTGAAGACATTTGGTGAAGTGTATATCCCTTTGTAATTTTGA 2760
Qy 6459 TACATTAATTTGATATTTGTTTATTTTGTGTAAGATGTAGAAATGTACTATGTTT 6518
Db 2761 TACATTAATTTGATATTTGTTTATTTTGTGTAAGATGTAGAAATGTACTATGTTT 2820
Qy 6519 ATGCTTCTACATCCAGTTTGTACAACTGGAATAATTAATTAATTAATTAATTAATTA 6568
Db 2821 ATGCTTCTACATCCAGTTTGTACAACTGGAATAATTAATTAATTAATTAATTAATTA 2870
RESULT 5
ABK94949
ID ABR94949 standard; cDNA: 6383 BP.
XX ABR94949;
AC 30-AUG-2002 (first entry)
DT 30-AUG-2002 (first entry)
XX 30-AUG-2002 (first entry)
DE Human novel polynucleotide #60.
KW Human: gene; ss: inflammatory condition; shock; sepsis; immune response;
cancer; wound healing; central nervous system disease; haematopoiesis;
peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW

KM bone degenerative disorder; periodontal disease; reperfusion injury;
KM lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KM allergic condition; thrombolysis; thrombosis; coagulation disorder;
KM fungal infection.
OS Homo sapiens.
PN WO200244340-A2.
XX 06-JUN-2002.
XX 30-NOV-2001: 2001WO-US47004.
XX 30-NOV-2000: 2000US-0028952.
XX (HISE-) HISEQ INC.
PI Tang YF, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI Yamazaki V, Ujwal ML, Drmanac RT;
XX MPI: 2002-508509/54.
DR P-PSDB: ABG66725.
PT Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing -
XX Claim 1: Page 469-477: 672pp: English.
XX The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human
CC novel polynucleotides of the invention.
XX
SQ Sequence 6383 BP: 1953 A; 1236 C; 1435 G; 1759 T; 0 other:
Query Match 24.2%: Score 1587: DB 24: Length 6383:
Best Local Similarity 66.0%: Pred. No. 0: Mismatches 1205: Indels 99: Gaps 12:
Matches 2519: Conservative 0:

DB 817 GGAAGCAGCAGTTTATCTGATATCTATTCAGGCTACGAGAGTGAGGTAGACATGTAGAT 876
QY 285 CTGATGGGTTGGCACAACACAGATGATTCGGAAGCAGCAGCATGANAAGACATT 344
DB 877 TTGACACGCTTTCAGAAGACCTGTTGATTCGAGCATACCAAGAAAGATGAAGAG 936
QY 345 GAGAGAGCATCAGATCCTCGATGAGCAGGACGATTGTGAGATGCGCCAGGAAGGAC 404
DB 937 ATTGATCGAACAGATCCATTGCGAGGGCCGAGATCTTGTGGAANTGCTTGAANAAGAA 986
QY 405 CCAATTGACCGAGACAGATGATGACATTTGAACAACCTTGGAAATTTATGACACAGTTGCC 464
DB 997 CCGCAGACAAAACTGATGATGACATTTGAACAATTCCTGAGATTATGACCAAGCTCC 1056
QY 465 GCTTTTCCCAATATGACATGTCATGAGGCGGAACCTGTCTGCTGATGATGTTGGCA 524
DB 1057 GCATTTGCAAAACATGACCATGTCTGTAGAGAGAACTGCTCAGTATGATGATTTGAA 1116
QY 525 GTGGTGAAGAGCAGGAGCCATGTTTAATGATGTGCAAGAGCTGCTCGTCA 584
DB 1117 GTGGTAGAGCAGGCTGAGCTATTTATTCGGAAGATGGCAAGAGCTTGCTCATGGTAT 1176
QY 585 GTGATTCATGATGATCTGTGAGAGTACTTATCCAGATGGAAGAGCAAAATACTGTGC 644
DB 1177 GTTATTTTAAAGCGCACTGTGGAATTCAGTCATCCAGATGGAAGATTTTCTTT 1236
QY 645 ATGGGAATATGTTTGGTGTCTCTCCTACATGACACAAAGATATCATGAAGAGTATG 704
DB 1237 ATGGAAATATGTTTGGAAATTTACTCCACCTGTGATAGCATATCATGATGAATGTTC 1296
QY 705 AGAACAAAGGTGATGATCTGCGAGTTTGTCTGATGCGCCAGCAAAATTTCTCCGAT 764
DB 1297 AGACATTAAGTATGATGTTGTGATGTTGTCTGATGCGCCAGCAAGATTATGAGATT 1356
QY 765 CTCATCAATGAGAAAAGAACATGCAAAAAGTTGAAGAGAGAGAGATTTGATAGTG 824
DB 1357 TTAACCATGTGGAAAAAAATACCCTAATGAAGTTGAGGAAGAGGAGAAATTTGATGTA 1416
QY 825 AAGAACACCGAAGAACTTGATGCACTGGAACAAGAAAGGACACATTTGATCAAGGT 884
DB 1417 CATGACATCTGGAACTAGACCGAGTGAACCAAGGAACACATTTGATCAAGGCA 1476
QY 885 ACCTGAAAGGTTAAACAATGCAATTTGGTGAAGAGATTCAGATGATTCACAACCTTC 944
DB 1477 ACACCTGAGCGCTCTCATTAATGCAATTTAATGAAGAACATTCATCGGATCCAACTTAT 1536
QY 945 ATGAAGACTTCTGTGACCTATGACCTTTCTTCTAGCCCAATGGAAGTGGCAAA 1004
DB 1537 ATGAAGATTTTCTATTAATTAAGTACAGACATTTCTTGAAGTCTTTGGATGTTGGATC 1596
QY 1005 AACTTATGAGTGTGTTTAATGACCCGAGCTCAGGATATAGTTTACACGGGTATATTA 1064
DB 1597 AAACATATTGGAATGTTTAAGATCGACAGCTTAAGATTAAGGTGACACGATTTGATTA 1656
QY 1065 TTGTGGTAAATATATCACTTCAATGATTTGAAGAGATCTCCAGATCATCATTTTGA 1124
DB 1657 TTATGGTAAATATATCAATTTAATGATTTGAAGGTACCTCTCATGATCCATTTTCTA 1716
QY 1125 GAAGAATTTGAAAACATCTGGAAGAGAGAAATGGGTGACACCTTAAGGCTTTGAAT 1184
DB 1717 GAGCAATTTGAAAATAATTCGGAAGATACAAAGATGAATGATCTCGCGGTATTGCAAT 1776
QY 1185 ATGCGGTGCTCTTAAGCAAAAAGAGATTGATGACGTTAAACAAAACATCCCGAGAA 1244
DB 1777 ATTGCGTGTGCTCAAAAGGCTAAGTGAGACAGATGTGCTGCAAAAAGCTTCCCGGAG 1836
QY 1245 GCTCCTTGGCTTTATCTTACTTGGAGGCTTGAGAGAGATTTGGAATCTTTGGTAC 1304
DB 1837 TCCCTCTCAATTCACCTTAATGAGGAGTGAAGAGGATTTGATATTTTGTGAA 1896
QY 1305 AGTGTAGATTCAAGTAGCAAGCAACTGAAGCAGCTTGGAAGCGGGGATCATGATATTA 1364
DB 1897 GAGTTGAACCTGTGCAAAAGCTGCTGATCAAGACTGAAGCTGAAGTGTGATCAGATTATG 1556

Qy	1365	GAGTAAATGGCCAAACACTTTTGAAAACATTTCAAGCTGTCAAAAGCGTAAATGGAAATCTTTGA	1422
Db	1957	GAACTAAATGGCAAAACCTTTGGAGAATATTACATTTATGAAAGCCGTTGAATTTTGAGG	2016
Qy	1425	AATTAACACGATTTATATCATCACTGTGAAGCAATATTTATTTGATTTTAAAGACTTCTA	1484
Db	2017	AATTAATACATCTTGCACTTACTGTGAAGCAACCAATTTTGTGTCCAAAGATTTACTT	2076
Qy	1485	ACAAGATTGTCAAGAGAGAAAGAAATGTTGCCCCCACTTCTTAAATTTGGTGACATT	1544
Db	2077	TTTAG---GACTGAACAGAGAAATCTGTGTCTCTCAATATTTCCAAAATTTCTGAAAAA	2133
Qy	1545	AAAAAGCCAGTGGCTACTCCATTCACAGATCTTGTGCTGATGTAGAACAGGTGATAGA	1604
Db	2134	AAAAAGT---AATGCCATTCTATCCAGCATGTGGCAGAGGATATTGAAACA---GACATCA	2187
Qy	1605	CTTGA AAAAGTGAACAAAAAGTAAAGCCCAACTGTG---GGAGGAAGAACAAAGCTG	1661
Db	2188	CAGGAGAAAGGAAGTAAAGAAAGTTAAGCAAAATACGTTTCAGTGGTCAGAAACAAAAATC	2247
Qy	1662	AAAAAGATCTGCACAAAGACTGGGATCAGTATTTCTTGGCCACAGAAACCATATGATGATT	1721
Db	2248	AGGAAATTTTGGATTAACACACGATTTAGATTTCTGCTCCCAAGCTATTTAGTGATGA	2307
Qy	1722	GGGATTGGTCAGTCAAGTACAGATAGTAAAGGATTAAGGCAGACAAAGACATCCCA	1781
Db	2308	GGCCTAAGCCAAATCACAAGATGACACATTTGTGGAAACAAAGGCATGTAGGCTATCTTG	2367
Qy	1782	ACTGCATTGGCTCTCAAGTGAACCTTATCATCCAGTAAATCTGATTTATTTGCAATCAAT	1841
Db	2368	GCTATATATGCCATCTCCCTGGAAACACTCTATCCAGACGCCCTGATCTCTCGACGCCATCC	2427
Qy	1842	CATGCATTTTACAGCTTCAGTGTACTCCGACTCTGGCAGATCAAGTCTAAAGGTTTTT	1901
Db	2428	ACCAAGTATGTGGATTTTTCATCTCTTCAGATATCCCTGATCAAGTATTAAGAGTTTTC	2487
Qy	1902	AAGCGTGAATGAGCAAAAGCCGCTACATCATCATGATAGTAAAGCACTACAGAAAGAAAGTG	1961
Db	2488	AAAGTGAATGAGCAAAAGTCTCTACATATATCATAGTAAAGACACACAGCTAAAGAAAGTA	2547
Qy	1962	GTCATTAGGCTATCAAGGAGTTTGGCTGTACTGCTACGCCACCCGGATCAATATTGACTATGT	2021
Db	2548	GTTTTATATGCTGTCTCAATGATTTTGGTTTGAACGGGTGCATCCGACATATTTCTCTGTG	2607
Qy	2022	GAGCTCTCTGCACACCTGAGGAGGAGTATCAAAACAAAGAAAGACTTCCAGATCAGCTTCC	2081
Db	2608	GAACTTTCTGTTACTCTCTGAGGGGTGTATAAACAGGAAGACTTCCAGATCACTTCTCC	2667
Qy	2082	AAACTTGCAGACAAATATCAACCTGAGTGGAAAGTATTTATCGAAAAACAATGGAACA	2141
Db	2668	AAATTTACTGTATAAATTTCAACTCACTCAATGGAAGGATTTACTTTAAAAAATTTACATGGAAACA	2727
Qy	2142	GAAACTCTTTTGTGATGATGATGATGCTCAGGAGTTGTTGAGAGAGCTCAAAATTTCCCTC	2201
Db	2728	GAAACCTTATGTTTCAGATGAAAGATGCTCAAGACATAGTTAAGGAAAGCCACACTATCCATG	2787
Qy	2202	CTTCAAGCTCAGACACTGTGGGAAGTTGGCAACACACTCTCTATGCGCAAAATTTGAACTCTT	2261
Db	2788	CTGCAAGCTCACTACCATTTGAGTGGGCCACCCAGCTGTCAATTAAGAGGACTTTGATTTGTTT	2847
Qy	2262	CGCAACATTTGAACCTTACGATATATAGATGATTTATTTAATCACTCAGATCAAAACAGC	2321
Db	2848	CGTATATTTGAACCGACTGACTACATGATCACTTTTAAAGTTAAATTTCCAAAAACAGA	2907
Qy	2322	TGTGCCAACTGGAAGAGATTTGAAGAGTCATTAACCGAAGAACTTTTGGGTAGCATCT	2381
Db	2908	AATACTCATTTGAAAGCGTTTGAAGACATTTTAAACCAAGAGACATTTCTGGGTTGCCCA	2967
Qy	2382	GAAATTTCCAGAGAAACAAACAGCTGGAAGAGATGAAGATCATTTAACAATTTCTCAG	2441
Db	2968	GAAATTTTAACTGAAGCAAAATCGCTCAACAGCATGAAGATTTATTAACATTTTATTTAA	3022

Qy	2442	AATGACAGTGCACGTAGGGATGCAAGAAATTTTAATCTCAATGTTTGCAATATCATGAGTGC	2501	
Db	3028	ATTGCACCTTCATTGTGCGAATAGTAAGAACTTCATTCATGTTTGGCAATTAATTAAGTGGC	3087	
Qy	2502	CTAAACCTGGCCACAGTGGCAGACTGGCAACGACCTGGGAGAAACTCCCAATTAATATC	2561	
Db	3088	TTGAACCTGGCACTGTGTAGCCAAAGACTCAGAGCACTTGGGAAAATTTCCAAAGCAAAATTC	3147	
Qy	2562	GA AAAA CTATTTCAAGATCTCCAAGACCTGTTTGATCTTCCGAAACATGAGCAAAATAT	2621	
Db	3148	GAGAAACATCTTCAAGATCTACAGACATTTTGTATCCATCTGAAACATGAGCAAAAGAT	3207	
Qy	2622	CGTAATGTTCTCAATGTGCAAAAATGTACAAACCTCCATATATCCCTATATCCCAATTATC	2681	
Db	3208	AGAAATATTTCTAGTAGTGAATGCAAGATATGACGCTCCCATATTAATTCACATCTTCCCTGTTC	3267	
Qy	2682	AAAAAGCATCTCACTCTTCCCTTCACGAAGAAATGACTCAAAAGTAGAGCGGCTGTCAAT	2741	
Db	3268	AAAAAGCATATGCAATTTCTACATGAAAGAAATGATCCAAAGTAGATGTTTAACTAATC	3327	
Qy	2742	TTTTGACAGCTTAAGGATGATGTGCAAAAAGAAATTCGTACAGTGGCGGATAGGCTTCAGTC	2801	
Db	3328	TTTGAAGAAGTTAAGATGATATTTCCAGAAATTCGCCAAGTTGTTCGAATGACTTCTGCT	3387	
Qy	2802	AATATGACCTCCCTCCCTCATGTTTCAGAGACTCGGAAMAAATATGGCGAGTTTGGGGTCT	2861	
Db	3388	AACATGGACCCCACTATGATGTT-----TCGACAGAGTCA 3423		
Qy	2862	CTGACCCAGGGTAGTACAAATGCAACAGTGTGATGTGTTCTCAGACAGTGGTCAATAA	2921	
Db	3424	CTGAGTCAAGGAAGCAACAAATTTAAACATGCTGGATGTT--TCAGGAGGTGCTCCAAAC	3480	
Qy	2922	AAGGGGTAGCTGTGATGTTCTTTCATGTCGCAAAAGCTTTATTAAGATGAGTGGCCAAATG	2981	
Db	3481	AAAAAGGCACGCGCAGACTCTGTGCTTAATGCCAAGAGCTATATGAGATGGCCAAATG	3540	
Qy	2982	GCTCGAAAGTGAAGCAGTACCTTTTCCAAATTTGAGCTAGAAATGACAGAGAGACTCT	3041	
Db	3541	GCAAGGAAGGTGAAGCAGTATCTTTCACAGCTCGATGTAGACAGATGAGGAAGTTC	3600	
Qy	3042	CAGACATTATCTCGCAGTGTGAGCGAGCAACCAACACATTTGGCTTAAGATTCCTGTGTAC	3101	
Db	3601	CAGATGATGTTCATTACAGTGGAGGACTGCAATATGATGCTTGGACCAAGATTTAAAGTGA	3660	
Qy	3102	AAAAAGCCTGTCAA--ATCCGACACTCTTCACSTACTCCAAAGGCGAGGTCACAAACG	3158	
Db	3661	AAAAAGTACAGCAATCATCTGAAATGTCTCCATGTGAGTCAAGTGTGCGCAAAACA	3720	
Qy	3159	AAAGCTCAGTCCCGCCACAGCCCCACAGAGACGACACACAGCACATTAATATCAACAG	3218	
Db	3721	ACTAAAGCCCACTTGATCAACCC-----CACACAGTAAGCGAC 3759		
Qy	3219	GGACTTACAGTTCCCGCGGTGCTCCCTTATCTTTCACGGAAGAAAGTGCCCGTAAAGGAT	3278	
Db	3760	GTGCTTACAGTGCAGAGCTGTTAATTTTGACCCCATCAGGAAGGGAACAACAAAAGAC	3819	
Qy	3279	CTCCCACTTTTGGATTAATCACTGTCCACAACTTTAAAAAAAATTTCTTCTTGCGAA	3338	
Db	3820	CTGTGAGTAATACAAAGTTTACCTGACAAAGTTTTAGSACAA-----CTGAA 3867		
Qy	3339	GAGGAACTTTTGAACGTACACAGAAACAGGCTGAAGATATCATATGATGACTTTCG	3398	
Db	3868	GAATAA-----GTGGTAAGAGCATACAGAGACATATTTCTGTGGGCTCATCT 3918		
Qy	3399	CAGCTTTTCTCTCTCTACTTCTTCCACAGAGTTCTCCAAAGAAAGGCTATACTTGGCT	3458	
Db	3919	TTACATTTCTAGTCCCTGCACTCTCTCAAGGCTCCCTCACAAAGGTTTACACTTATTT	3978	
Qy	3459	CCCATGTGAGCTGTGATAATTTTCAAGATCTGTGTCACAGTGAATTTCTTCACGATCC	3518	
Db	3979	CCATCAGCTAATCTGACAACTGTCTGACTGCAGCCCATATGAGATTTCTTCACGGTCC	4038	
Qy	3519	AGTATTTTAGCAATTCGTTCTTTGACTCAGTGCACAGTCTCATCTGACAGATGAGAGGCGC	3578	

Db 4039 AGCATGTGAGCAATGTTCTGTGTGACATGCTCTGACAGCTCTACAGATGAAACGCTGT 4098
3579 CAGAGCATTTGCTGTGACATGCTGTGAAACAACTAGGGATGGCGAGAGAGCGG 3638
4099 TCCTCTCAGGCCCCGTGGACGTCCCTGTAATCAGCTGGGGCATTTGGAAAGACAGACGCT 4158
3639 ACCATGATGAACCTGATCAGTATACCTTGGGCTCTATGACCAATGTGCCAGGCGCGA 3698
4159 TCAGGATGAGGATCATATGTCACATGCGCCCTGGGTGGACACTCTTGAAGCATCTCTA 4218
3699 GCGTTATATGCTACAGCTACAGTATTTCTTCCAGACAGACAGAGAACTTCCAGAT 3758
4219 ATCAAGTGTGTAGCTGTCTCA-----TCGCTGTGAGCAATGAAGAGATTCTCAAGAG 4272
3759 CAGGGGATCGCGCGCTGATGCTGCTGACAGTGGCGCTGGAGCTGGAGAGTATGC 3818
4273 CATATCATTT-----ATGAAAGCAGCTGACAGAGTGTGTGAAGTTGAGCTTGTGT 4323
3819 TCAGTGTGCTCCCATGATATATATACAGAGATCCAGACAGAGAGAGCTGGGA 3871
4324 TCAGAGAGCTCCCATGATCAACTTCCAAAGCTTCCAAACCCAAAAGCTGGGA 4376

RESULT 6

AACT6867
ID AACT6867 standard; cDNA: 6403 BP.

XX AACT6867;

DT 08-FEB-2001 (first entry)

DE Human ORF2422 polynucleotide sequence SEQ ID NO:4843.

XX
Human: open reading frame; ORF: detection; cytosolic; hepatotropic;
vulnery; antiparalytic; antiparkinsonian; nootropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
antianemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antinflammatory disease; coagulation;
thrombosis; contraceptive; ss.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CUBAGEN CORP.

XX Shinketsu RA, Leach M;

XX WPI: 2000-602362/57.

XX P-PSDB: AAB42658.

XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -

PS Claim 5; Page 4024-4028; 5507pp; English.
XX
XX AACT4446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytosolic; hepatotropic; vulnery;
CC antiparalytic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORF-associated disorder. The
CC nucleic acids can be used to express ORF proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease, to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 6403 BP; 1973 A; 1237 C; 1434 G; 1757 T; 2 other;

XX Query Match 24.1%; Score 1582.4; DB 21; Length 6403;

XX Best Local Similarity 65.9%; Pred. No. 0; Mismatches 1206; Indels 102; Gaps 12;

XX Matches 2528; Conservative 0;

XX 45 AGCATGTTTCTTCATATGAAACCACTAGCAATCCAGCTAACCATGATTATGGC 104

XX 637 AGAAGATTTTCGAAATTAATTAAGAGAGAGCGCAAACTTACTGATGTGGAG 696

XX 105 CAGCAGAGAAACATCTACTCTGCAAGATTGCAAAACTGCATCTTACTGACGCTC 164

XX 697 GTTAAAGCATTTCTTCTCCAGCTGATCTTACCAAGATCATCTCAAGAAACCTT 756

XX 165 CACCACAGGTGACCGACGTTCTTCTAGCCATGAGATGATATCATGATGATCTT 224

XX 757 CATCCAGGTGATCTATGCTCTTCTAGCACTGTGTTGCAATGCACTGACTCTT 816

XX 225 GAGAGAGAGAGTCTTCTGATATCTACAGCGACAGAAAGCAGGCTGTTATGAGC 284

XX 817 GGAAGACAGAGTTTATCTGATATCTTACAGCTACGAGAGTGAAGATGATGAT 876

XX 285 CTGAGTGGGTTGCCAAGACAGAGTGGATTCCGAAGACGACGACGATGAAGACATT 344

XX 877 TTGACAGGCTTCCAGAAAGAGCCTGTTGATTTGAGGATGACGAAGAGATGAAG 936

XX 345 GAGAGAGACATGATCTCTGATGACAGGACATGTTGAGACATGCTCAGAGAGAC 404

XX 937 ATTGATCGAAGACATCTATGACAGGCGAGATCTGTTGGAATGCTTTGAAAGAA 996

XX 405 CCAATTGACCGAGACAGATGATGACATTTGAACAACCTCTTGGAATTTAGCACAGTTCCT 464

XX 997 CCGACAGACAAACATGATGATGACATTTGACATTTGCTGAGATTTAGCACAGCTCCT 1056

XX 465 GCTTTTCCATATGACATGTCGAGGCGGAACTCTGCTGATGATGCTTGGCA 524

XX 1057 GCATTTTCCAAACATGACCAATGCTGTAAAGAGAGAACTGCTCAGTGAATTTTGA 1116

XX 525 GTGGTGAAGAGACGAGGACATGATGTTAAATGATGTTGAAGCTGAGCTCCTGCTCA 584

XX 1117 GTGTGAGAGAGGCTGAGCTTATTTCTTGAAGATGGGCAAGAGCTGCTCATGCTAT 1176

XX 585 GTGATTTCTCAATGATGATGAGAGTACTTATTCAGATGGAAGCAAGCAATACTGTC 644

XX 1177 GTTATTTTAAAGGCACTGTGGAATTCAGTATCCAGATGGAAGTGAATTTGTTT 1236

XX 645 ATGGGAATAGTTTGTGCTCTCTTACCATGACAAAGATATCATGAAGAGATGATG 704

XX 1237 ATGGGAATAGTTTGTGATTAATCTCCACTCTGATTAAGCAGATGATGATGAATTTGTC 1296

705 AGAACAAGGTGGATGCTCCAGTTTGTCTGATAGCCACAGATTACTGCCGATTT 764
1297 AGGCTAACTACATGATTTCTGATGATGCTGATGAGCCAGCAAGATTATGGAGAAAT 1356
765 CTCATCAAGTAGAAGAAACATGCAAAAAGTTGAAGAGAGAGAAAGATTGTTAGGTC 824
1357 TTAAACCATGTGAAAAAATACCCTAAAGTTGAGAGAAAGAGAAATGTTATGTA 1416
825 AAAGAACGAGAGAACTTGATCGAAGTGAACAAGAGGAGACATTTGTCATCAAGGCT 884
1417 CATAGCATGCGAGATAGACCGGAGTGAACGAGAAAGACATTTGATGATAGGCA 1476
885 ACCTCAGAAAGGTTAACATGATTTTGTGGAAGAGATTCAGTAGATTCACATTC 944
1477 ACACCTGAGCCTCTCATATGTCATTTAATAGAAAGAACATTCATCGGATCCACCTTA 1536
945 ATAGAGACTTTCTGTTGACCTATAGACTTTTCTTCTAGCCCAATGGAAGTGGGCAAA 1004
1537 ATAGAGATTTTCTATTACTTACAGGACATTTCTTGAAGCTTTGATGTTGGGATC 1596
1005 AAGTTATTGAGGTGTTTATGACCCAGCCTCAGGATAGTTTACACGGGTAGATTTA 1064
1597 AACCTATTGGAATGTTTAAATGACAGCTTAAAGATAGGTAGACAGGATTTGATTA 1656
1065 TTGTGGTAAATCACTGATGATGATGATGAGAGATCTCTGCAATGATCGATTTTAA 1124
1657 TTATGGGTAAATCACTGATTTTATGATTTTGAAGTGACCTGCTGATAGCTCGATTTCTA 1716
1125 GAAGATTGGAACAACTCGAAGAGAGAAAGATGAGTGAACACATCCGAGAA 1184
1717 GAGGATTTGAAAAAATCTGGAAGATACAAAGATGATGATCTCCGTTGATGAT 1776
1185 ATGCGCTGCTCTCAAGCAAAAGAGATGATGATGATGATGATGATGATGATGATGAT 1244
1777 ATGCTGCTGCTCAAGCAAAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1836
1245 GCT 1304
1837 TCCCT 1896
1305 AGCTGATTTCAAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1364
1897 GAGGTGAACCTGTGAGCAAGCTGCTGATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1956
1365 GAAGTAATGCGCAAAACCTTTGAAACATTCAGCTGCAAAAGCTATGGAATTTCTTGA 1424
1957 GAAGTAATGCGCAAAACCTTTGAAACATTCAGCTGCAAAAGCTATGGAATTTCTTGA 2016
1425 AATTAACACATTTATCTATCACTGGAAGAACCAATTTATTTGATTTAAAGCACTTCTA 1484
2017 AATTAATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2076
1485 ACAGATTTGTCAGAGAGAGAAAGATGAGTGGCCCACTTCTTAAATTTGCTGACAT 1544
2077 TTTTGG---GACTGAGCAAGAGAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2133
1545 AAAAAGCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1604
2134 AAAAG---TAATCGCCATTTCTATCCACATGTCGAGAGATTTGTAACA---GACATCA 2187
1605 CTTGAAAAGTGAACAAAAAGTAAGCAACACTGCTG---GAGAGAGAGAAACAGCTG 1661
2188 CAGGAGAAAGAGATGAGAAAGTTAAAGCAAAATCTGTTTCAAGTGAAGAAACAAATC 2247
1662 AAAAAGATCTGACAGACACTGGAATGATCTTGGCAGAGAAACCATATCATATGATTT 1721
2248 AGGAAGATTTTGGATTTAAACACAGATTTAGTATCTTGGCTCCAAAGCTATTTAGTAGGA 2307
1722 GGGATGTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1781
2308 GGCCTAAGCCATATCAAGATGACAGCTTTGTGGAGAACAGCACTGTAGGCAATAGTCTG 2367

1782 ACTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1841
2368 GCTATTAATGCCCAATCCCTGGAAACCTCTCAWCCAGAGCCCTGATCTCTGCAAGCCTTAC 2427
1842 CATGCAATTTTACCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1901
2428 ACCGATATGTTGGATTTTCCAAATCTTCAAGATATCTCTGATCATATTTAAAGATTTTC 2487
1902 AAGGCTATCGAAGAGCCCTATCATGATGATGATGATGATGATGATGATGATGATGATGAT 1961
2488 AAAGTGAATGAGCAAAAGTTGCTTACATATATCATAGTAAAGACCAACAGCTGTAAGAGTA 2547
1962 GTCATTCAGCTATCAGGAGATTTGCTGTTATCTGCAACCCCGGATCAATATCATATGAT 2021
2548 GTTTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2607
2022 GAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2081
2608 GAAATTTCTGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2667
2082 AAACCTGCAAGCAATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2141
2668 AAATTAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2727
2142 GAAACTTTTGTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2201
2728 GAACTTTATTTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2787
2202 CTTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2261
2788 CTTGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2847
2262 CGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2321
2848 CGTATATTTTCAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2907
2322 TGTGCCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2381
2908 AATATCTATTTTGAAGGTTTGAAGCTGATGATGATGATGATGATGATGATGATGATGAT 2967
2382 GAAATTTCTGAGAAACAAACAGCTGAGAGATGATGATGATGATGATGATGATGATGATGAT 2441
2968 GAAATTTTAACTGAGCAAAATCAAGCTGATGATGATGATGATGATGATGATGATGATGAT 3027
2442 ATAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2501
3028 ATTCACCTTATTTGTCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3087
2502 CTAACCTGACCAAGTGAAGCTGCAAGCAAGCTGGAAGAACTTCCCAATTAATGAC 2561
3088 TTGAACCTGAGATCTGATGAGCAAGCTGCAAGCAAGCTGGAAGAACTTCCCAAGCAATGAC 3147
2562 GAAAACTATTTCAAGATTTCTCAAGACCTGTTGATCTTCCGAAACATGCGCAAAATAT 2621
3148 GAGAAACATTTCTCAAGATTTCTCAAGACATTTTGAATCCATCTGAAACATGGAAGATAT 3207
2622 CGTATTTCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2681
3208 AGAATATTTCTTGAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3267
2682 AAAAAGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2741
3268 AAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3327
2742 TTTGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2801
3328 TTTGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3387
2802 AACATGAGACCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2861
3388 AACATGAGACCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3423
2862 CTCAGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2921

```

Db 3424 CTGAGTCAAGGAAGCAACAATTCAACATGCTGATGT---TCAGGAGAGTGTCTCAACAA 3480
OY 2922 AACGGGTACGTGTGTAGTCTTCTTCAATGCCAAAAGCTTTATGACATGCCCAATG 2981
Db 3481 AAAAGGGCAGCGCGAGCTCTCTGTTAAATGCCAAGAGTATATGAGTGCCTCAAAATG 3540
OY 2982 GCTGSAAGAGTGAAGCAGTACCTTTCATTTTGGAGTGAAGTGAAGCGAGAGTCTT 3041
Db 3541 GCAGAGAGAGTGAAGCAGTATCTTTCCAGTCTGATGTGAGACAGATGAGGAGAACTTC 3600
OY 3042 CAGACATTATCTCTGAGTGTGAGCCAGCAACACACATTTGCTTAAGATCCGTGTGAC 3101
Db 3601 CAGATGATGATCATATACAGTGGAGAGCTGCATATGTATGATACCTTGACCAAGATTAAGTGAG 3660
OY 3102 AAAAACCCTGTCA-----ATCCGAGACCTCTCCAGTACTCCAAAGGCGAGGTCACAA 3155
Db 3661 AAAAGATCAGCCCAAGNNATCATCTGAATGTCTCAGTGTGAGTGTGAGTGTGAGTGTGAG 3720
OY 3156 CAGAAAGCTCAGTCCCTGCACAGCCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3215
Db 3721 ACACCTAAGGCCACCTTGACATCAACC-----CACAGAGTAAAGC 3759
OY 3216 CAGGAGCTACAGGTTCCCGCGCTGCTTATCTTCAAGAGAAAGTGCCTGAAG 3275
Db 3760 CAGGTGCTTACAGTGGCAGAGCTGTAAATTTGACCCCATGAGAAAGGACAAACAA 3819
OY 3276 GATCTCCACCTTTTGGCATAACTCTCCCAAGCTTTTAAAAAATCTCTTGTGTCT 3335
Db 3820 GACCTGACATGATACAACTTCTCAGAAAGTTTGAAGCAAA-----CT 3867
OY 3336 GAGAGAGAAAGTTTGAAGCTCAGAAACAGAGCTGAAGATCAATATCAATATGATCT 3395
Db 3868 GAAGAAATTA-----GTGTAAGAACCATACAGAAACATCTTTTGTGGGCTCA 3318
OY 3396 TCCAGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3455
Db 3919 TCTTTCATCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3978
OY 3456 GCTCCAGAGTGTACTGTGATTAATTTTTCAGATTTGTGTGACAGTGAATTTTCTTCA 3515
Db 3979 ATTCCATCACTCAATCTGACAACTGTCTGACTCCAGCATGATGATGATTTCTTCA 4038
OY 3516 TCCAGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3575
Db 4039 TCCAGCATCTGAGCAATTTTCTGTTGACTCTGATGCTGACCTTACAGATGAGCG 4098
OY 3576 CCGCAGAGGCAATCTGTGACATCTGTGAAACAAACCTAGGATGGGACAGATGAGAGG 3635
Db 4099 TGTTCCTCTCAGGCGCTGGGAGTCCGTAATCCAGTGGGCAATTTGAAAGACAGAGCAC 4158
OY 3636 CGACCATGATTTAACCCTGATGATAGCTTGGGTCTTATGACCATATGTCAGAGGCG 3695
Db 4159 GCTTCAGGAGTAGAGATCTAGTCAACATGAGCCGTGGGTGAGACACTCTTCCAGCCATCT 4218
OY 3696 CGAGGCTTATATGCTACAGTACAGTAAATTTCTTCCAGCAACAGAGAACTTTCCAG 3755
Db 4219 CTAAATCAAGTGTAGTGTCTCA-----TCGTCTGTGAGCAATGAGGATTTCTCAA 4272
OY 3756 GATCAGGAGGAGTGGCGGCTCACTTGTATGCTGTGACAGTGTGGGAGGAGCTGAGCTCA 3815
Db 4273 GAGCATATCATTT-----ATAGAAGCAGGTGAGAGTGTGTGGAATTTGAGACTTCG 4323
OY 3816 TGTCTCAAGTGTCTCCATGATTAATATACAGACATCCAGCAGCAGAGAGCTGGGA 3871
Db 4324 TGTTCAGCAGCTTCATGACAACTTCCAAAGCTTCCAAAGCCAAAGCTGGGA 4379

```

RESULT 7
 AAL20515
 ID AAL20515 standard; cDNA; 481 BP.
 AC XX
 AAL20515;

```

XX 07-DEC-2001 (first entry)
DT Human breast cancer expressed polynucleotide 12972.
DE Human; breast cancer; cell marker; cytosolic; ss.
KW Homo sapiens.
OS WO200151628-A2.
PN 19-JUL-2001.
PD 10-JAN-2001; 2001MO-US00798.
PE 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
PX (MIL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA Lillie J, Xu Y, Wang Y, Steinmann K.
PI WPI; 2001-451856/48.
PT New peptide useful as a marker for the diagnosis of breast cancer
PS Claim 1; Page 2298; 3695pp; English.
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterizing treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity.
SO Sequence 481 BP; 141 A; 136 C; 96 G; 108 T; 0 other;
Query Match 7.2%; Score 471; DB 22; Length 481;
Best Local Similarity 100.0%; Pred. No. 3.1e-108;
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2999 GTACCTTTCGAATTTGGAGCTAGAAATGAGAGAGAGTCTTACGATTAATCTGCA 3058
Db 9 GTACCTTTCGAATTTGGAGCTAGAAATGAGAGAGAGTCTTACGATTAATCTGCA 68
OY 3059 GTGTGAGCCAGCAACCAACCATTTGCTTGAAGATCTGTGACAAAAGCTGTCAATC 3118
Db 69 GTGTGAGCCAGCAACCAACCATTTGCTTGAAGATCTGTGACAAAAGCTGTCAATC 128
OY 3119 CGAGACCTTTCAGTACCTTCAAGGCGAGGCTCAACACAGAAAGCTGCTCCGCCACA 3178
Db 129 CGAGACCTTTCAGTACCTTCAAGGCGAGGCTCAACACAGAAAGCTGCTCCGCCACA 188
OY 3179 GCGCCAGCAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3238
Db 189 GCGCCAGCAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 248
OY 3239 GTCCCTTTATCTTCAAGAGAAAGTCCGTAAGATCTCCACCTTTGGCATAAA 3298
Db 249 GTCCCTTTATCTTCAAGAGAAAGTCCGTAAGATCTCCACCTTTGGCATAAA 308
OY 3299 CTCTCCACAAGCTTTAAAAAATTTCTTCTTGTCTGAAGAGGAAGTTGGAACGTCA 3358
Db 309 CTCTCCACAAGCTTTAAAAAATTTCTTCTTGTCTGAAGAGGAAGTTGGAACGTCA 368

```

```

QY 3359 CAGAAACAGCTGAGATACATATCAATGATCTTGCAGCTTCTTCTCTCTAC 3418
      |||
Db 369 CAGAAACAGCTGAGATACATATCAATGATCTTGCAGCTTCTTCTCTCTAC 428
      |||
QY 3419 TTCTCCACAGAGTTCTCCAGAAAGGCTATCTTGGCTCCAGGCTAC 3469
      |||
Db 429 TTCTCCACAGAGTTCTCCAGAAAGGCTATCTTGGCTCCAGGCTAC 479
      |||

RESULT 8
AAL1617
ID AAL1617 standard; cDNA: 467 BP.
XX
AC AAL1617;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 4074.
XX
KM Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PE 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192089.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steilmann K;
XX
DR WPI: 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1: Page 744; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 467 BP; 139 A; 131 C; 93 G; 104 T; 0 other:
Query Match 7.0%; Score 459; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.3e-105; Indels 0; Gaps 0;
Matches 459; Conservative 0; Mismatches 0;

QY 2999 GTACCTTCCAAATTTGAGCTAGAAATGACGAGAGCTTTCAGACATTTATCTCTGCA 3058
      |||
Db 9 GTACCTTTCAAATTTGAGCTAGAAATGACGAGAGCTTTCAGACATTTATCTCTGCA 68
      |||
QY 3059 GTGTGAGCGAGCAACCAACATTTGCTTAAGAAATCTGTGACAAAACCTGTCAAAATC 3118
      |||
Db 69 GTGTGAGCGAGCAACCAACATTTGCTTAAGAAATCTGTGACAAAACCTGTCAAAATC 128
      |||
QY 3119 CGAGACCTCTCCAGTACTCAAGGCGAGGTACAAAGAAAGCTCCCTGCGACA 3178
      |||

```

```

Db 129 CGAGACCTCTCCAGTACTCAAGGCGAGGTACAAAGAAAGCTCCCTGCGACA 188
      |||
QY 3179 GCCCAGAGAGAGCCACCACACATAAATACACAGGAGTACAGTCCGCGCT 3238
      |||
Db 189 GCCCAGAGAGAGCCACCACACATAAATACACAGGAGTACAGTCCGCGCT 248
      |||
QY 3239 GTCCCTTATCCCTTCAGGAGAAAGTCCGTAAGAGATCTCCACCTTTGGCATAAA 3298
      |||
Db 249 GTCCCTTATCCCTTCAGGAGAAAGTCCGTAAGAGATCTCCACCTTTGGCATAAA 308
      |||
QY 3299 CTCTCCACAAGCTTTAAAAAATTTCTTTGTCTGACAGAGAGATTGGAAAGCTCA 3358
      |||
Db 309 CTCTCCACAAGCTTTAAAAAATTTCTTTGTCTGACAGAGAAATTGGAAAGCTCA 368
      |||
QY 3359 CAGAAACAGGCTGGAAGATACAAATATCAATGATCTTGCAGGCTTCTCTCTCTAC 3418
      |||
Db 369 CAGAAACAGGCTGGAAGATACAAATATCAATGATCTTGCAGGCTTCTCTCTCTAC 428
      |||
QY 3419 TTCTCCACAGAGTTCTCCAGAAAGGCTATCTTGGC 3457
      |||
Db 429 TTCTCCACAGAGTTCTCCAGAAAGGCTATCTTGGC 467
      |||

RESULT 9
ABK34853/c
ID ABK34853 standard; cDNA: 4298 BP.
XX
AC ABK34853;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA for novel secreted protein, SEQ ID 622.
XX
PR Human; ss; gene; secreted protein; immune deficiency; viral infection;
PR bacterial infection; fungal infection; autoimmune disorder; burn;
PR rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
PR diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
PR Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
PR coagulation disorder; haemophilia; inflammatory disorder; ulcer;
PR tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
PR lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
PN WO200177290-A2.
XX
PD 18-OCT-2001.
XX
PR 29-MAR-2001; 2001WO-US10295.
XX
PR 06-APR-2000; 2000US-194941P.
XX
PA (GENY ) GENETICS INST INC.
XX
PI Wong GC, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gutukota K, Graham JR;
XX
DR WPI: 2002-179323/23.
XX
PT Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT
XX
PS Claim 1: Page 336-337; 339pp; English.
XX
CC The invention relates to 625 polynucleotides which have been derived from
CC a variety of human tissue sources and sequences that hybridise to them.
CC proteins, their complements and sequences that hybridise to them.
CC Also included are a vector comprising the polynucleotide, a host cell
CC transformed with the vector, the proteins encoded by the
CC polynucleotides, antibodies that bind to the proteins and identification
CC of modulators of the proteins or the expression of the polynucleotide.

```


CC The polynucleotides can be used as probes for the identification
CC and isolation of full length cDNA and genomic DNA. The polynucleotides
CC and proteins can also be used as nutritional supplements. The protein
CC is useful in the treatment of various immune deficiencies and disorders
CC such as viral infections, bacterial infections, fungal infections,
CC autoimmune disorders (e.g. Rheumatoid arthritis, multiple sclerosis,
CC conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment
CC of burns, incisions and ulcers. The proteins are also useful for
CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.

SQ Sequence 4298 BP; 1258 A; 925 C; 941 G; 1174 T; 0 other;

Query Match 6.2%; Score 404.8; DB 24; Length 4298;
Best Local Similarity 99.5%; Pred. No. 4.7e-91;
Matches 406; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1710 TACATGATATGGAGTGGTGCAGTCTCAAGATGACAGCATAGTATAGGACAGACA 1769
DB 3548 TTCAGTATATTGGATGGTGCAGTCTCAAGATGACAGCATAGTATAGGACAGACA 3489
QY 1770 AAGCATATCCCAACTGCTGCTGTCAGTGCAGACCTTATCATCCGTAATCTGATTTA 1829
DB 3488 AACCATATCCCAACTGCTGCTGTCAGTGCAGACCTTATCATCCGTAATCTGATTTA 3429
QY 1830 TTGCAGTCAATCATGCTGCTGCTGTCAGTGCAGACCTTATCATCCGTAATCTGATTTA 1889
DB 3428 TTGCAGTCAATCATGCTGCTGCTGTCAGTGCAGACCTTATCATCCGTAATCTGATTTA 3369
QY 1890 CTAAAGGTTTTTAAGGCTGATCAGCAAGCCGCTACATCATGATGATGATGAGACACTACA 1949
DB 3368 CTAAAGGTTTTTAAGGCTGATCAGCAAGCCGCTACATCATGATGATGATGAGACACTACA 3309
QY 1950 GCAAAGGAAGTGTCTATCAGGCTATCAGGAGTTTGTCTTACCTCCACCCGGATCAA 2009
DB 3308 GCAAAGGAAGTGTCTATCAGGCTATCAGGAGTTTGTCTTACCTCCACCCGGATCAA 3249
QY 2010 TATTCATATGAGGCTCTGTCTACACCTGAGGAGTATCAAAAGAAAGACTTCCA 2069
DB 3248 TATTCATATGAGGCTCTGTCTGTACACCTGAGGAGTATCAAAAGAAAGACTTCCA 3189
QY 2070 GATCAGCTTCCAACTTGCAGACAGATACAACTGATGGAAGGTAT 2117
DB 3188 GATCAGCTTCCAACTTGCAGACAGATACAACTGATGGAAGGTAT 3141

RESULT 10
ABL38290/c
ID ABL38290 standard; cDNA; 402 BP.

XX ABL38290;

DT 08-APR-2002 (first entry)

DE Human colon tumour antigen polynucleotide SEQ ID NO:1879.

XX Human: colon cancer; colon tumour antigen; cytostatic; vaccine;

KW colon tumour metastatic antigen; diagnosis; gene; ss.

XX Homo sapiens.

OS WO200196388-A2.

XX 20-DEC-2001.

XX 08-JUN-2001; 2001WO-US18557.

PR 09-JUN-2000; 2000US-210899P.
PR 20-FEB-2001; 2001US-270216P.
XX
XX
PA (CORI-) CORIXA CORP.

PI Jiang Y, Harlocker SL, Secrist H;

DR WPI: 2002-114514/15.

PT Novel isolated colon tumor polynucleotide differentially expressed in
PT colon tumor or colon metastatic tumor and polypeptides encoded by them,
PT useful for inhibiting development of cancer in patient -

PS Claim 1: SEQ ID 1879; 105pp; English.

CC ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
CC which were isolated from human colon tumour and colon metastatic tumour
CC cDNA libraries. (I) have cytostatic activity and can be used in vaccine
CC production. (I) can be used for stimulating and/or expanding T cells
CC specific for a tumour protein on contact with the T cells. They are also
CC useful for inhibiting the development of cancer in a patient. (I) can be
CC used as probes or primers for nucleic acid hybridisation, for preparing
CC mutant species primers, or primers for use in genetic constructions. (I)
CC can be used in the diagnosis of a colon tumour.

SQ Sequence 402 BP; 150 A; 67 C; 62 G; 121 T; 2 other;

Query Match 6.1%; Score 399.2; DB 24; Length 402;
Best Local Similarity 99.8%; Pred. No. 3.5e-90;
Matches 399; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5960 TACTCAAGTGTGAGGAAGCTGCTGCTGATTTTTCATTAATTCAGTGCATCAT 6019
DB 400 TACTCAAGTGTGAGGAAGCTGCTGCTGATTTTTCATTAATTCAGTGCATCAT 341
QY 6020 ATTGATCAGTATGATTAACGTAATAGCTTCAATTTTAAAGTGAATGCACTGTTTT 6079
DB 340 ATTGATCAGTATGATTAACGTAATAGCTTCAATTTTAAAGTGAATGCACTGTTTT 281
QY 6080 TCACGTATCAACAATGTCAGTCTTATTTATTAATCTCTTCTGATCATGCAATTT 6139
DB 280 TCACGTATCAACAATGTCAGTCTTATTTATTAATCTCTTCTGATCATGCAATTT 221
QY 6140 GTCTACTTGTCTTATTCATTTGATTAATGATTTGATTAATTTACATGTAATGCAATTA 6199
DB 220 GTCTACTTGTCTTATTCATTTGATTAATGATTTGATTAATTTACATGTAATGCAATTA 161
QY 6200 TTTCGCAAGTTTTTATATAGGCTATGACCTCATGTGATATAGAAAGACAGAAATCTA 6259
DB 160 TTTCGCAAGTTTTTATATAGGCTATGACCTCATGTGATATAGAAAGACAGAAATCTA 101
QY 6260 GCTCTACCAACAAGTGCACAAATGTTATCTAGCATTAAAGTATAGACATAGAGCT 6319
DB 100 GCTCTACCAACAAGTGCACAAATGTTATCTAAGCATTAAAGTATAGACATAGAGCT 41
QY 6320 GCTAATCTCAGTTCGCTGTGATGTCAGTGCAGATGT 6359
DB 40 GCTAATCTCAGTTCGCTGTGATGTCAGTGCAGATGT 1

RESULT 11
AAA59384
ID AAA59384 standard; DNA; 801 BP.

XX AAA59384;

DT 07-NOV-2000 (first entry)

DE DNA clone encoding a murine guanine nucleotide releasing factor 4.

XX guanine nucleotide releasing factor 4; GRF-4; Ras activator;

KW Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;
cell differentiation; cell transformation; neuronal disorder; ss.

```

XX OS Mus sp.
XX FH Key Location/Qualifiers
XX CDS 1..801
FT FT /*tag- a
FT FT "contains 5 internal stop codons and encodes
FT FT AAB07793; no termination codon given"
FT CDS 2..799
FT FT /*tag- b
FT FT "contains 8 internal stop codons and encodes
FT FT AAB07794; no termination codon given"
FT CDS 3..800
FT FT /*tag- c
FT FT "contains 5 internal stop codons and encodes
FT FT AAB07795; no termination codon given"
XX PN WO20043510-A2.
XX PD 27-JUL-2000.
XX PE 20-JAN-2000; 2000WO-CA00042.
XX PR 20-JAN-1999; 99CA-2259830.
XX PA (HSCR-) HSC RES & DEV LP.
XX PI Rolin D, Pham N;
XX DR WPI: 2000-499228/44.
XX DR P-PSDB; AAB07793, AAB07794, AAB07795.
XX PT Nucleic acids encoding guanine nucleotide releasing factor-4 useful for
XX PT the treatment of cancers and neuronal disorders.
XX PS Disclosure: Flg 19B; 89pp; English.
XX CC The present sequence encodes a murine guanine nucleotide releasing
XX CC factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both
XX CC in vitro and in vivo. It directly binds cyclic adenosine monophosphate
XX CC (cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cAMP)
XX CC binding domain). GRF4 directly connects cAMP-generating (e.g. G protein
XX CC coupled receptors) or cAMP-generating pathways to Ras. GRF4 activates
XX CC Ras in response to elevation of intracellular cAMP and/or cAMP. GRF4
XX CC is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of
XX CC the Ras signalling pathway controls numerous cellular functions, such as
XX CC cell metabolism, proliferation, differentiation and transformation.
XX CC Therefore modulation of Ras activity may provide a mechanism for
XX CC controlling diseases. GRF4 polynucleotides and polypeptides may be used
XX CC in the treatment of diseases associated with inappropriate GRF4
XX CC expression and activity such as cancers and neuronal disorders. The
XX CC GRF4 polypeptides may be used as antigens in the production of
XX CC antibodies against GRF4 and in assays to identify modulators (agonists
XX CC and antagonists) of GRF4 expression and activity. The anti-GRF4
XX CC antibodies and GRF4 antagonists may also be used to down regulate GRF4
XX CC expression and activity. Inhibition of Ras can reduce celluiose
XX CC proliferation and cancers.
XX SQ Sequence 801 BP; 173 A; 266 C; 232 G; 130 T; 0 other;
XX
XX Query Match 5.9%; Score 389.2; DB 21; Length 801;
XX Best Local Similarity 74.6%; Pred. No. 1.7e-87;
XX Matches 586; Conservative 0; Mismatches 183; Indels 17; Gaps 7;
QY 4034 AACCTGGGCTCTTCCACAGGTTACTGGGAGAGACTAGAAAGTGCACAGCGCAAT 4093
Db 15 AACCTGAGCTCCACCGCGTGGCGCCGCTCTAGAACTAGTGATGCCCGGCTGCAG 74
QY 4094 AAGCGGAGGGGTGAAGATGTTTCATTGAAGCCGAAGAGTACCTTAACGCTGTGT 4153
Db 75 GAATTCAAGCGGTGGAGAGATGTCTCCGTGAGGAGAGAGACAGATGATGCCCT 134
QY 4154 GACTACGAGAGAAACCAAGCTGTCCCATGCTGCCACATAGCTGTGGCATCAAGTAC 4213

```

```

Db 135 GACTACGAGAGAGAGCAAAACCTGTCTCCATTCCTGCCACATAGCTGTGACCGCAGAC 194
QY 4214 TACAAGGGGCTCATTTGCACGAAAGAGGCGAGTATTCGAGAGACCCCGCCACCTCC 4273
Db 195 TACCAAGGAGCTCATCGCACGAAAGAGGAGGAGTACCGGGAGCGGCTCCACCTCC 254
QY 4274 CGGCTACATTGGAAATTTCCATTACGACTTTCCAGAAAGGGACCTCCATCCAGCAGAA 4333
Db 255 AGGCTACGTGGGCATCCCATTTGCCATTTCCAGAAAGGGCTTTGCCACCGGCAAGAA 314
QY 4334 ACCGCGGAGCTACAAAGTGGGCTTTCAGAGATCGGGATGGTGCAGATCTTCGACAC 4393
Db 315 GCCCGCGATTTACAAAGTGGGCTTTCAGAGATCGGGATGGTGCAGAGGCGCCACTGAGC 374
QY 4394 AGCTGGGCTTTCATCCGTAACAGACAGCCATGGGATATCCCAACGACGACGAGCGTGA 4453
Db 375 CCCGGCACC-----GGGCAAGACCGCGCTGACCGCCAGCCAGCCGCGGCGAG 425
QY 4454 CAACCTCAGTGCATTAACCGAAGAGTGCACCGCGCTCGCCCTTATACGTCGA 4513
Db 426 CAAGCCACAGTGGCACAGACCCAGGACGACAGACCCAGCGCTCGGCCCTTCAAGCGCA 485
QY 4514 AGGGTTTCCA--CCGAGAGAGATGAAGTGAACAAGTTTCTGCTGTTTGAGCAGAC 4571
Db 486 GGCTTCGAGAGAGCGGAGAGAGAGAGTAAGTAACAAGTGTCTGTTTGAGCGCAGGC 545
QY 4572 TTTTCTGGAAGCAGAGCGGCGCCAGTGAAGAGCAGCAAGACAGCTCTGAGCATTTG 4631
Db 546 --TCTTGTATCCACAGTGAAGCAGC--CAAGAGAGACAGCAAGACAGCTCTGAGCATTTG 602
QY 4632 GAGCCTTGAACATCAGATTTCTGAGAGAGGTGAGCAGATTTGCTCTCTCTCTTAA 4691
Db 603 GAGCCTTGGACAGCACA--TCTGAGATGCTGAGCAGATTTGCTCTCTCTCTCTTAA 660
QY 4692 AGCAGCATGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4751
Db 661 AGCAGCATGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
QY 4752 GAATTTGCCCTGACACTTTTTCAGATTTGTTGAATGACAGTGCAGCATCTTCG 4811
Db 721 GAATTTGCCCTGACACTTTTTCAGATTTGTTGAATGACAGTGCAGCATCTTCG 779
QY 4812 AGCTCC 4817
Db 780 AGCTGC 785

```

```

RESULT 12
ABL37607/c
ID ABL37607 standard; cDNA; 403 BP.
XX ABL37607;
AC XX
DT 08-APR-2002 (first entry)
XX
XX Human colon tumour antigen polynucleotide SEQ ID NO:1196.
DE
DE Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
KW colon tumour metastatic antigen; diagnosis; gene; ss.
XX
XX Homo sapiens.
OS
OS WO200196388-A2.
XX
XX 20-DEC-2001.
XX
XX 08-JUN-2001; 2001WO-us18557.
XX
XX 09-JUN-2000; 2000US-210899P.
XX
XX 20-FEB-2001; 2001US-270216P.
XX
XX (CORI-) CORIXA CORP.

```


XX Jiang Y, Harlocker SL, Secretist H;
XX
XX MPI: 2002-114514/15.
XX
XX Novel isolated colon tumor polynucleotide differentially expressed in
XX colon tumor or colon metastatic tumor and polypeptides encoded by them,
XX useful for inhibiting development of cancer in patient -
XX
XX Claim 1: SEQ ID 1196; 105pp; English.
XX
XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (1)
XX which were isolated from human colon tumour and colon metastatic tumour
XX cDNA libraries. (1) have cytostatic activity and can be used in vaccine
XX production. (1) can be used for stimulating and/or expanding T cells
XX specific for a tumour protein in contact with the T cells. They are also
XX useful for inhibiting the development of cancer in a patient. (1) can be
XX used as probes or primers for nucleic acid hybridisation, for preparing
XX mutant species primers, or primers for use in genetic constructions. (1)
XX can be used in the diagnosis of a colon tumour.
XX
XX Sequence 403 BP; 151 A; 67 C; 62 G; 121 T; 2 other:
XX
XX Query Match 5.9%; Score 388.2; DB 24; Length 403;
XX Best Local Similarity 99.5%; Pred. No. 2.1e-87;
XX Matches 399; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
XX
XX 5960 TACTCAAAATGCGAGAACTGCTGCTGATTTTTCATTAAATTCAGCTGATCAT 6019
XX :|||||
XX 401 HACTCAATGTGAGAACTGCTGCTGATTTTTCATTAAATTCAGCTGATCAT 342
XX
XX 6020 ATTGATCAGTAAAGCTAAATAGCTCAATTTTAAAGTGAATGCACTG-TTT 6078
XX :|||||
XX 341 ATTGATCAGTAAAGCTAAATAGCTCAATTTTAAAGTGAATGCACTGTTT 282
XX :|||||
XX 6079 TTCACTGTAAACAATGTCAGTCTTTAATTAATTTCTCTGATTCATGGCATT 6138
XX :|||||
XX 281 TTCACTGTAAACAATGTCAGTCTTTAATTAATTTCTCTGATTCATGGCATT 222
XX :|||||
XX 6139 TGGTACTGCTTATTCATTTGCAATTTGCAATTTTGTATTTTACATGTAATGATT 6198
XX :|||||
XX 221 TGGTACTGCTTATTCATTTGCAATTTGCAATTTTGTATTTTACATGTAATGATT 162
XX :|||||
XX 6199 ATTGTCAGTTTATTTATATAGCTATGACCTCATGTCATATAGAAAGCAAAATCT 6258
XX :|||||
XX 161 ATTGTCAGTTTATTTATATAGCTATGACCTCATGTCATATAGAAAGCAAAATCT 102
XX :|||||
XX 6259 AGCTTACCAACAAGTTCACAAATGTTATTAAGCATTAAGTAATGTAACATAGAC 6318
XX :|||||
XX 101 AGCTTACCAACAAGTTCACAAATGTTATTAAGCATTAAGTAATGTAACATAGAC 42
XX :|||||
XX 6319 TGGTAAATTCAGTTGCTGCTGATGTCAGTCAAGTCAAGAAATG 6359
XX :|||||
XX 41 TGGTAAATTCAGTTGCTGCTGATGTCAGTCAAGTCAAGAAATG 1
XX
XX RESULT 13
XX ABL15285
XX ID ABL15285 standard; cDNA; 4656 BP.
XX AC ABL15285;
XX XX
XX 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 40337.
XX XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX XX
XX PN WO200171042-A2.
XX XX

PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX MPI: 2001-656860/75.
XX
XX P-PSDB: ABB71182.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1: SEQ ID NO 40337; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL101840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4656 BP; 1131 A; 1338 C; 1259 G; 928 T; 0 other:
XX
XX Query Match 4.4%; Score 289; DB 23; Length 4656;
XX Best Local Similarity 58.1%; Pred. No. 7.3e-62;
XX Matches 530; Conservative 0; Mismatches 380; Indels 3; Gaps 1;
XX
XX 1869 CMTGACTGGCAGATCAAGTCTAAGGCTTTTAAGCTATACGAAAGCCGTACATC 1928
XX :|||||
XX 2260 CCCGATTACCCGAGCCAGCTGCTCAAGTGTACAGCAACAGAGCTCAAGTATGTA 2319
XX :|||||
XX 1929 ATGATCAGTAAAGCACTTAAGCAAGAAAGTGTCTAATAGCTATACAGAGATTGCT 1988
XX :|||||
XX 2320 CTGATCTACAGAGAGACAGACACCCAGAGAGTGTATGCTGACGCTGACAGATTGGA 2379
XX :|||||
XX 1989 GTTACTGCCACCCCGATCAATATCTATGTGAGGCTCTGTGACACCTGAGAGGTA 2048
XX :|||||
XX 2380 ATACACGACCCAGCTGCAATTTCTGCTATGCAAGTGAAGTGTGGCAGCGGGGTATG 2439
XX :|||||
XX 2049 ATCAANCAAGAGACTTCAGATCAAGTCTTCCAACTTGCAGACAAATATCAACTAGT 2108
XX :|||||
XX 2440 GTAAAGCAGCGCGGTGGCCGATCACTGCAAGAACTGGCCGCAATTAAGCTTTGCG 2499
XX :|||||
XX 2109 GGAAGTATTATCTGAAAACACATGGAACAAACTCTTTGTTAGATGAATGATCT 2168
XX :|||||
XX 2500 GCGGTTACTACCTCAGATTAAGATAGACGCAACCGCTGGTTCAGACGAAGTGGCC 2559
XX :|||||
XX 2169 CAGAGTGTGTTAGAGAGAGTCAAAATTTCCCTCTTCAAGCTCAGACAGTGGAAATGCA 2228
XX :|||||
XX 2560 CTGGAACGTGTAGCGCAGTGAACGTGCACTCTGCACTGAAATGATAGAGCTGGCC 2619
XX :|||||
XX 2229 ACAAGCTCTATGCGAAATTTTGAACCTTTCGCAACATTAAGCTACTGAATATATA 2288
XX :|||||
XX 2620 ATTCAAGCTCACTACAGAGACTTTGCGAATTTGCGCAAGATGAGTCCACCGAGATAGTG 2679
XX :|||||
XX 2289 GATGATTTATTTAACTCAGATCAAAAACACAGCTGTGCCAAGCTGAGAGATTTGAAGA 2348
XX :|||||
XX 2680 GAGCAATTTGTCAGTTGCGCTCAGCTAGGAGAGTGCATGCTGACAGCAAGTTTGGCAA 2739
XX :|||||
XX 2349 GTCATTAAACAGAAACATTTTGGGTAGCATCTGAATTTCTCAGAGAAACAAACAGCTG 2408
XX :|||||
XX 2740 CTGCTCAATCGTAAAGTGTGGTGTGATGATGATTTGGCAGAGCAACATGTTGTG 2799
XX :|||||

BLANK PAGE

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 18, 2003, 08:55:54 ; Search time 15987 Seconds
(without alignments)
11956.416 Million cell updates/sec

```
Title: US-09-911-826A-1
Perfect score: 6568
Sequence: 1 ctgcacatcgtgagagattg.....aaataataataatacat 6568
```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

```
Searched:      2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 41092800
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

Database :

1:	gb_ba:	*
2:	gb_hhg:	*
3:	gb_in:	*
4:	gb_om:	*
5:	gb_ov:	*
6:	gb_pac:	*
7:	gb_ph:	*
8:	gb_pl:	*
9:	gb_pr:	*
10:	gb_ro:	*
11:	gb_srs:	*
12:	gb_un:	*
13:	gb_un:	*
14:	gb_vl:	*
15:	em_ba:	*
16:	em_fun:	*
17:	em_hum:	*
18:	em_in:	*
19:	em_mu:	*
20:	em_om:	*
21:	em_or:	*
22:	em_ov:	*
23:	em_pat:	*
24:	em_pl:	*
25:	em_pl:	*
26:	em_ro:	*
27:	em_srs:	*
28:	em_un:	*
29:	em_un:	*
30:	em_hhg_hum:	*
31:	em_hhg_in:	*
32:	em_hhg_other:	*
33:	em_hhg_mus:	*
34:	em_hhg_pln:	*
35:	em_hhg_prod:	*
36:	em_hhg_nam:	*
37:	em_hhg_vrt:	*
38:	em_hhg:	*
39:	em_hhg_hum:	*
40:	em_hhg_mus:	*
41:	em_hhg_other:	*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	6568	100.0	6568	6	AX028128	AX028128 Sequence
2	6568	100.0	6568	6	AX408888	AX408888 Sequence
3	6568	100.0	6568	6	AB002311	AB002311 Human mRNA
4	2011	30.6	160185	9	AC074344	AC074344 Homo sapi
5	1612	24.5	5742	9	AF394782	AF394782 Homo sapi
6	1585.4	24.1	6390	9	AF478567	AF478567 Homo sapi
7	1584.4	24.1	6401	9	AF117947	AF117947 Homo sapi
8	1581.4	24.0	1803	9	AF070570	AF070570 Homo sapi
9	1577.6	24.0	4806	9	AF478468	AF478468 Homo sapi
10	1502.6	22.9	4176	9	AF478466	AF478466 Homo sapi
11	1348.6	20.5	181796	2	AC124358	AC124358 Mus muscu
12	1341.4	20.4	3450	2	AK074318	AK074318 Homo sapi
13	1290	19.6	185108	2	AC118835	AC118835 Rattus no
14	863.4	13.1	194590	2	AL645962	AL645962 Mus muscu
15	845.6	12.9	148466	2	AC023167	AC023167 Mus muscu
16	728.6	11.1	195471	2	AC124186	AC124186 Mus muscu
17	716.2	10.9	173519	2	AC130840	AC130840 Mus muscu
18	697.2	10.6	3596	10	BC025553	BC025553 Mus muscu
19	601.4	9.2	2200	9	AK056296	AK056296 Homo sapi
20	560.8	8.5	51067	2	AC100578	AC100578 Mus muscu
21	399.2	6.1	402	6	AX341632	AX341632 Sequence
22	390.4	6.0	799	6	AX028130	AX028130 Sequence
23	390.4	5.9	4111	3	AF308449	AF308449 Caenorhab
24	390.4	5.9	4413	3	AF170796	AF170796 Caenorhab
25	380.4	5.9	4413	3	AF308447	AF308447 Caenorhab
26	380.4	5.9	4550	3	AF308448	AF308448 Caenorhab
27	368.2	5.9	403	6	AX340949	AX340949 Sequence
28	364.4	5.1	185108	2	AC118835	AC118835 Rattus no
29	333.2	5.1	181796	2	AC124358	AC124358 Mus muscu
30	326.4	5.0	3670	9	HSMB00982	HSMB00982 Homo sapi
31	289	4.4	76375	2	AC017591	AC017591 Drosophill
32	289	4.4	209071	3	AC005285	AC005285 Drosophill
33	289	4.4	214622	3	AC093100	AC093100 Drosophill
34	289	4.4	261872	3	AE003613	AE003613 Drosophill
35	287.4	4.4	5025	3	AF434686	AF434686 Drosophill
36	230.8	3.5	192912	10	AL607091	AL607091 Mouse DNA
37	227.8	3.5	263	6	AX155654	AX155654 Sequence
38	219.6	3.3	85181	9	AC004622	AC004622 Homo sapi
39	219.4	3.3	108424	10	AL669827	AL669827 Mouse DNA
40	208.2	3.2	38799	3	CEK04D7	Z6964 Caenorhabdi
41	198	3.0	2675	10	BC019702	BC019702 Rattus no
42	174.4	2.7	110523	2	AC103244	AC103244 Rattus no
43	152.2	2.3	158	6	AX072819	AX072819 Sequence
44	150	2.3	133255	2	AX072819	AX072819 Rattus no
45	139.2	2.1	51067	2	AC100578	AC100578 Mus muscu

ALIGNMENTS

RESULT 1				
AX028128				
LOCUS	AX028128	6568 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0043510.			
ACCESSION	AX028128			
VERSION	AX028128.1	GI:10188937		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 6568)			
AUTHORS	Rotin,D. and Pham,N.			
TITLE	Ras activator nucleic acid molecules, polypeptides and methods of use			

|||||
Db 1501 AAAAAAGAAATGTGCCCCACCTCTCTAAATGGTGACATTTAAAAAGGCGAGCGCT 1560
QY 1561 ACTCCATTCCAGATCTTGGTGTAGATGTGAACAGGTGTAGAGCTTGA AAAAGTGAACA 1620
Db 1561 ACTCATTCACATCTTGGTGTAGATGTGAACAGGTGTAGAGCTTGA AAAAGTGAACA 1620
QY 1621 AAAAAAGTAAACCAACACTGTGGGAGAGAACAGCTGAAAAAGATCTCGACACA 1680
Db 1621 AAAAAAGTAAACCAACACTGTGGGAGAGAACAGCTGAAAAAGATCTCGACACA 1680
QY 1681 CTGGATCAGTATCTTGGCCACAGAAACCATACATGATATGGGATTTGCTCAGTCAAG 1740
Db 1681 CTGGATCAGTATCTTGGCCACAGAAACCATACATGATATGGGATTTGCTCAGTCAAG 1740
QY 1741 ATGACAGCATAGTATTAAGGACAGAACACATCCCACTGCATGGCTGTCAAG 1800
Db 1741 ATGACAGCATAGTATTAAGGACAGAACACATCCCACTGCATGGCTGTCAAG 1800
QY 1801 GAACCTTATCATCAGTATCTGATTTATGTCAGTCACATCGCATTTTATAGACTCA 1860
Db 1801 GAACCTTATCATCAGTATCTGATTTATGTCAGTCACATCGCATTTTATAGACTCA 1860
QY 1861 GTGCTACTCTGACTTGCAGATCAAGTCTAAGGGTTTTAAGGCTGATCAGCAAAAGCC 1920
Db 1861 GTGCTACTCTGACTTGCAGATCAAGTCTAAGGGTTTTAAGGCTGATCAGCAAAAGCC 1920
QY 1921 GCTACATCATGATCATAGTACAGACACTACAGCAAGAAAGTGTGATTCAGGCTATAGGG 1980
Db 1921 GCTACATCATGATCATAGTACAGACACTACAGCAAGAAAGTGTGATTCAGGCTATAGGG 1980
QY 1981 AGTTTGTCTGTACTGCGACCCCGATCAATATTCACTATGTAGGCTGTGTACACCTG 2040
Db 1981 AGTTTGTCTGTACTGCGACCCCGATCAATATTCACTATGTAGGCTGTGTACACCTG 2040
QY 2041 AGGAGTAACTCAACAAAGAAAGCTTCCAGATCAGCTTTCCAAACTTGCAGACAGAAATAC 2100
Db 2041 AGGAGTAACTCAACAAAGAAAGCTTCCAGATCAGCTTTCCAAACTTGCAGACAGAAATAC 2100
QY 2101 AACTGAGTGAAGGTATTTATGAAAAACACATGGAACAGAAACTCTTTGTTCAGATG 2160
Db 2101 AACTGAGTGAAGGTATTTATGAAAAACACATGGAACAGAAACTCTTTGTTCAGATG 2160
QY 2161 AAGATCTCAGAGATGTTGAGAGAGATCAAAATTTCCCTCCTTACGCTCAGACACTGTGG 2220
Db 2161 AAGATCTCAGAGATGTTGAGAGAGATCAAAATTTCCCTCCTTACGCTCAGACACTGTGG 2220
QY 2221 AAGTTGCAACACAGCTCTATGCGAAATTTGAACTTTGCAACATTTGAACCTACTG 2280
Db 2221 AAGTTGCAACACAGCTCTATGCGAAATTTGAACTTTGCAACATTTGAACCTACTG 2280
QY 2281 AATATATAGATGATTTATTTAAACAGATCAAAAACCGCTGTGCAACCTGAAGAGAT 2340
Db 2281 AATATATAGATGATTTATTTAAACAGATCAAAAACCGCTGTGCAACCTGAAGAGAT 2340
QY 2341 TTGAAGAAGTCAATTAACAGAAACATTTTGGGTAGCATCTGAAATTTCCAGGAACA 2400
Db 2341 TTGAAGAAGTCAATTAACAGAAACATTTTGGGTAGCATCTGAAATTTCCAGGAACA 2400
QY 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTCAATCAAGATAGCATGTAGGG 2460
Db 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTCAATCAAGATAGCATGTAGGG 2460
QY 2461 AATGCAAGAATTTTAACTCAATGTTTGAATCATCGTGGCCCTAAACCTGGGACCCAGTGG 2520
Db 2461 AATGCAAGAATTTTAACTCAATGTTTGAATCATCGTGGCCCTAAACCTGGGACCCAGTGG 2520
QY 2521 CAAGACTGGGAAGGAGCTGGGAGAACTTCCCAATAAATACGAAAAACTATTTCAAGATC 2580
Db 2521 CAAGACTGGGAAGGAGCTGGGAGAACTTCCCAATAAATACGAAAAACTATTTCAAGATC 2580
QY 2581 TCCAGACCTGTTTGAATCTTCCAGAAACATGCAAAATATCGTAAATGTTCTCAATAGTC 2640
Db 2581 TCCAGACCTGTTTGAATCTTCCAGAAACATGCAAAATATCGTAAATGTTCTCAATAGTC 2640
|||||
Db 2581 TCCAGACCTGTTTGAATCTTCCAGAAACATGCAAAATATCGTAAATGTTCTCAATAGTC 2640
QY 2641 AAAATATACACCTCCCATATTCCTCTATTCGCAATATCAAAAAGATCTCAGCTTCC 2700
Db 2641 AAAATATACACCTCCCATATTCCTCTATTCGCAATATCAAAAAGATCTCAGCTTCC 2700
QY 2701 TTCCAGAGAAATGACTCAAAAGTAGACGGGCTGTCAATTTTGAAGCTTAAGATGA 2760
Db 2701 TTCCAGAGAAATGACTCAAAAGTAGACGGGCTGTCAATTTTGAAGCTTAAGATGA 2760
QY 2761 TTGCAAAAGAAATTCGTACGTTGGCCGAATGGCTTCAGTGAACATGACCCCTGCTCA 2820
Db 2761 TTGCAAAAGAAATTCGTACGTTGGCCGAATGGCTTCAGTGAACATGACCCCTGCTCA 2820
QY 2821 TGTTCAGACCTGGGAGAGAAATGGCGGAGTTTGGGGTCTCTACCCAGGGTAGTACA 2880
Db 2821 TGTTCAGACCTGGGAGAGAAATGGCGGAGTTTGGGGTCTCTACCCAGGGTAGTACA 2880
QY 2881 ATGCAACAGTGTAGATGTTGCTCAGACAGTGTGTCATAAAAAGCGGTPACGTAGTT 2940
Db 2881 ATGCAACAGTGTAGATGTTGCTCAGACAGTGTGTCATAAAAAGCGGTPACGTAGTT 2940
QY 2941 CCTTCTCAATGCGCAAAAAGCTTTATGAAGATGCCAAATGGCTCGAAAAAGTGAACAGT 3000
Db 2941 CCTTCTCAATGCGCAAAAAGCTTTATGAAGATGCCAAATGGCTCGAAAAAGTGAACAGT 3000
QY 3001 ACCTTTCCAATTTGGAGCTAGAAATGGAGAGAGAGTCTTCAGACATTTATCTCTGAGT 3060
Db 3001 ACCTTTCCAATTTGGAGCTAGAAATGGAGAGAGAGTCTTCAGACATTTATCTCTGAGT 3060
QY 3061 GTGAGCCAGCAACCAACATTTGCTTAAGAATCTGTGTGCAAAAAGCTGTAAATCCG 3120
Db 3061 GTGAGCCAGCAACCAACATTTGCTTAAGAATCTGTGTGCAAAAAGCTGTAAATCCG 3120
QY 3121 AGACCTCTCAGTAGTCCAAAGGGCAGGGTCAACAGAAAGCTCAGTCCCTCCACAGC 3180
Db 3121 AGACCTCTCAGTAGTCCAAAGGGCAGGGTCAACAGAAAGCTCAGTCCCTCCACAGC 3180
QY 3181 CCCAGCAGCAGCCACACAGCATTAATAATCAACAGGAGACTACAGGTTCCGCCGTGT 3240
Db 3181 CCCAGCAGCAGCCACACAGCATTAATAATCAACAGGAGACTACAGGTTCCGCCGTGT 3240
QY 3241 CCCATTATCCTTACAGGAGAAAGTCCCGTAAGAGATCCCAACCTTTTGGCATAACT 3300
Db 3241 CCCATTATCCTTACAGGAGAAAGTCCCGTAAGAGATCCCAACCTTTTGGCATAACT 3300
QY 3301 CTCACAGCTTTAAAAAAATTTCTTTCTTGTCTGAAGAGAAAGTTTGAACGTACACA 3360
Db 3301 CTCACAGCTTTAAAAAAATTTCTTTCTTGTCTGAAGAGAAAGTTTGAACGTACACA 3360
QY 3361 AGAAACAGGCTGAAGATACAAATTCAAATGCAATCTTCCGAGCTTCTCCCTCACTT 3420
Db 3361 AGAAACAGGCTGAAGATACAAATTCAAATGCAATCTTCCGAGCTTCTCCCTCACTT 3420
QY 3421 CTCACAGAGTCTCCAAAGAAAGCTATACTTTGGCTCCCACTGTACTGTGATTAAT 3480
Db 3421 CTCACAGAGTCTCCAAAGAAAGCTATACTTTGGCTCCCACTGTACTGTGATTAAT 3480
QY 3481 TTTCAAGATCTGTCAAGATGAATTTCTTACGATCAGTATTTTGAAGCTTGTGCTT 3540
Db 3481 TTTCAAGATCTGTCAAGATGAATTTCTTACGATCAGTATTTTGAAGCTTGTGCTT 3540
QY 3541 TTGACACAGTGGCAGTCACTCAGATGAGAGGGCCGAGAGCAATCTGTGATGATCG 3600
Db 3541 TTGACACAGTGGCAGTCACTCAGATGAGAGGGCCGAGAGCAATCTGTGATGATCG 3600
QY 3601 TTGAAACAAACCTAGGAGATGGCAGATGAGAGGGGACATGATTTGAACCTGATCAGT 3660
Db 3601 TTGAAACAAACCTAGGAGATGGCAGATGAGAGGGGACATGATTTGAACCTGATCAGT 3660
QY 3661 ATAGCTTTGGGGTCTATGCAACAATGTCCGAGGGCCGAGGCTTATATGCTACAGCTACAG 3720
Db 3661 ATAGCTTTGGGGTCTATGCAACAATGTCCGAGGGCCGAGGCTTATATGCTACAGCTACAG 3720
|||||

QY	4801	AGCAATCTTGAGCTCCACACTGTGTGCTGCTCCACATCAACAGATATCAATTC	4860
Db	4801	AGCAATCTTGAGCTCCACACTGTGTGCTGCTCCACATCAACAGATATCAATTC	4860
QY	4861	CAAGATCATCAACAAGATGATTCACCTGCGCTGCATCTTCAATCCGCGAAGATT	4920
Db	4861	CAAGATCATCAACAAGATGATTCACCTGCGCTGCATCTTCAATCCGCGAAGATT	4920
QY	4921	TTTTTAATCTTCCTTTAGATTTCAAATGCTGACATGCTGACATTTGATTTGGATATG	4980
Db	4921	TTTTTAATCTTCCTTTAGATTTCAAATGCTGACATGCTGACATTTGATTTGGATATG	4980
QY	4981	AGAAAAGCTAGCCATTGAACTACTGTGGGCGTTTAAACCAACAAGAAAGAAAA	5040
Db	4981	AGAAAAGCTAGCCATTGAACTACTGTGGGCGTTTAAACCAACAAGAAAGAAAA	5040
QY	5041	CAATGAATCTCTTGAGTACAGATGCTGTGCACCTGTTTACATGCTCCTTTAAAAA	5100
Db	5041	CAATGAATCTCTTGAGTACAGATGCTGTGTGCACCTGTTTACATGCTCCTTTAAAAA	5100
QY	5101	AAAAATATGATTTAAATATTTGTTGACAGAGAAAAATATATCCATTATATGATTCAG	5160
Db	5101	AAAAATATGATTTAAATATTTGTTGACAGAGAAAAATATATCCATTATATGATTCAG	5160
QY	5161	TTATTTTAAACCTTTAGTAGAGGGTGGCCAGCGCTGGTTTCTGAAAAACCAATATGCCGG	5220
Db	5161	TTATTTTAAACCTTTAGTAGAGGGTGGCCAGCGCTGGTTTCTGAAAAACCAATATGCCGG	5220
QY	5221	ACAGGGGTGGCCACACCAAGAAAGCGGGAACCTGCGTGTGACCTGGCGTCCCATG	5280
Db	5221	ACAGGGGTGGCCACACCAAGAAAGCGGGAACCTGCGTGTGACCTGGCGTCCCATG	5280
QY	5281	TCCTTCTGGTCTCACCCGCGAAGGCCCTATCTCGAATATGAATATTTAGCCAAATTAA	5340
Db	5281	TCCTTCTGGTCTCACCCGCGAAGGCCCTATCTCGAATATGAATATTTAGCCAAATTAA	5340
QY	5341	TACCAAGACACCTCATCTGCTCTTCCCACTGATGGGGTCTTCTGTAAACCTGTTTG	5400
Db	5341	TACCAAGACACCTCATCTGCTCTTCCCACTGATGGGGTCTTCTGTAAACCTGTTTG	5400
QY	5401	CACATGGCCAGGGGAGGAACTAGACACCTGTGCTCTGACCTTTAGAGAGCGAGG	5460
Db	5401	CACATGGCCAGGGGAGGAACTAGACACCTGTGCTCTGACCTTTAGAGAGCGAGG	5460
QY	5461	ACGGTGTCAATGGCGGATGTGTCTGCTGCATTTAGATGATGCAAAACCCCATTTTTAA	5520
Db	5461	ACGGTGTCAATGGCGGATGTGTCTGCTGCATTTAGATGATGCAAAACCCCATTTTTAA	5520
QY	5521	GTATATATTTCTTGAATTTTGTATTTAGATGATGATGATGATTTGTTTGTTTTGT	5580
Db	5521	GTATATATTTCTTGAATTTTGTATTTAGATGATGATGATGATTTGTTTGTTTTGT	5580
QY	5581	TTTTTTTTAAGAAACATTTATAACTGGATGCAATTCAGTGAAGAAAGCAGCTTGGAGATT	5640
Db	5581	TTTTTTTTAAGAAACATTTATAACTGGATGCAATTCAGTGAAGAAAGCAGCTTGGAGATT	5640
QY	5641	GGAGCTAATGCGCAGCTGTTTACTGCTCTTTCAAGACAGCGCTCCCTTATGATTTAGG	5700
Db	5641	GGAGCTAATGCGCAGCTGTTTACTGCTCTTTCAAGACAGCGCTCCCTTATGATTTAGG	5700
QY	5701	ATTAGGGAATTAACAGCCTTTAAACCTGTATATAAAGATCAAAAACCTGGTTAGACATGCC	5760
Db	5701	ATTAGGGAATTAACAGCCTTTAAACCTGTATATAAAGATCAAAAACCTGGTTAGACATGCC	5760
QY	5761	AGCCTTTGCAAGGCAAGTTAGTCAACCAAGACTAACCTCCAACTGGCTTTATGAGCGCTG	5820
Db	5761	AGCCTTTGCAAGGCAAGTTAGTCAACCAAGACTAACCTCCAACTGGCTTTATGAGCGCTG	5820
QY	5821	CATATGACAAGGCTAATAGTAGAACAACATCTGCTACAGCTGCTATTAACCTATAT	5880
Db	5821	CATATGACAAGGCTAATAGTAGAACAACATCTGCTACAGCTGCTATTAACCTATAT	5880
QY	5881	GACTGAAATGAGCCCTCCACTCTATTTTGTGTGTGTTGACAGACTCCGGAAGATGA	5940


```
|||||
Db 5881 GACGTAAATGACCCCTCCACCTATTTTGTGTGTTTGGACACACACGACCCGAAAGATGA 5940
Oy 5941 AGCGTCCCAATCTGAGTGTACTCAATGTGAGGACTGCTGCTTGGATTTTTTTC 6000
Db 5941 AGCGTCCCAATCTGAGTGTACTCAATGTGAGGACTGCTGCTTGGATTTTTTTC 6000
Oy 6001 ATTAATTCAGCTGATCATATGATGATGATTAACGTAATTAATTAATTAATTA 6060
Db 6001 ATTAATTCAGCTGATCATATGATGATGATTAACGTAATTAATTAATTAATTA 6060
Oy 6061 GTGGAATTCAGCTGTTTTTTCAGTATCAAAACATGTCAGTCTTATTTAATTAATTC 6120
Db 6061 GTGGAATTCAGCTGTTTTTTCAGTATCAAAACATGTCAGTCTTATTTAATTC 6120
Oy 6121 CTTCGTATCATGCGATTTGCTTACTGCTTATTTACATTTGCAATTAATGATTT 6180
Db 6121 CTTCGTATCATGCGATTTGCTTACTGCTTATTTACATTTGCAATTAATGATTT 6180
Oy 6181 TTACATGTAATATGATTAATTTGCGAGTTTATTTATATGAGTATGACCTGATG 6240
Db 6181 TTACATGTAATATGATTAATTTGCGAGTTTATTTATATGAGTATGACCTGATG 6240
Oy 6241 ATGAAAGACAGAAATCTAGCTCTACACAGTTCACAAATGTTATCTTAACATTAAGT 6300
Db 6241 ATGAAAGACAGAAATCTAGCTCTACACAGTTCACAAATGTTATCTTAACATTAAGT 6300
Oy 6301 AATTGTAGAACATAGGACTGCTAATCTCAGTTCGCTCTGTGATGTCAGTCAAGTGA 6360
Db 6301 AATTGTAGAACATAGGACTGCTAATCTCAGTTCGCTCTGTGATGTCAGTCAAGTGA 6360
Oy 6361 CAATTAACGAGTATTCCTCCTACTTGTGATGATGATGATGATGATGATGATGATG 6420
Db 6361 CAATTAACGAGTATTCCTCCTACTTGTGATGATGATGATGATGATGATGATGATG 6420
Oy 6421 AGACATTTGCTGAGTCTGATCCCTTTTGTATTTTAAATCAATTAATGATGATGATG 6480
Db 6421 AGACATTTGCTGAGTCTGATCCCTTTTGTATTTTAAATCAATTAATGATGATG 6480
Oy 6481 TATATTTTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 6540
Db 6481 TATATTTTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 6540
Oy 6541 CAAGCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6568
Db 6541 CAAGCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6568

RESULT 2
AX408888 6568 bp DNA linear PAT 14-JUN-2002
LOCUS AX408888
DEFINITION Sequence 1535 from Patent W00229103.
ACCESSION AX408888
VERSION AX408888.1 GI:21441593
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 1535 11-Apr-2002;
FEATURES
source
1. 6568
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. AB002311"
BASE COUNT 1974 a 1400 c 1463 g 1731 t
ORIGIN
Query Match 100.0%; Score 6568; DB 6; Length 6568;
```

```
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 CTTCGCATCTGAGAGATTTGATCATGATGCTAAATTCAGTTCACGATATGTTCTTCA 60
Db 1 CTTCGCATCTGAGAGATTTGATCATGATGCTAAATTCAGTTCACGATATGTTCTTCA 60
Oy 61 TTATGAACCTATCAATCCAGCTAACCATGAGTTTGGGCCGAGGAGAAACACT 120
Db 61 TTATGAACCTATCAATCCAGCTAACCATGAGTTTGGGCCGAGGAGAAACACT 120
Oy 121 CACTTCCTGACATTTTCACAAATGCTATCTAGTACAGTCTCCACCCAGGTGACC 180
Db 121 CACTTCCTGACATTTTCACAAATGCTATCTAGTACAGTCTCCACCCAGGTGACC 180
Oy 181 ACGTTTCTTACAGCATTTGAGATGATGATGATGATGATGATGATGATGATGATG 240
Db 181 ACGTTTCTTACAGCATTTGAGATGATGATGATGATGATGATGATGATGATGATG 240
Oy 241 CGATATCTACAGGCGCAGAAAGCGAGGCTGTGATATGACCTGATGGTTGCCAG 300
Db 241 CGATATCTACAGGCGCAGAAAGCGAGGCTGTGATATGACCTGATGGTTGCCAG 300
Oy 301 AAACAGCAGTGTATCCGAGACAGACAGATGAAGACATTTGAGAGATCAGATC 360
Db 301 AAACAGCAGTGTATCCGAGACAGACAGATGAAGACATTTGAGAGATCAGATC 360
Oy 361 CTCTGATGAGAGGACATTTGTGAGAGACTGCTAGAGAGACCAATTTGACCGGACAG 420
Db 361 CTCTGATGAGAGGACATTTGTGAGAGACTGCTAGAGAGACCAATTTGACCGGACAG 420
Oy 421 ATGATGACATTTGAACACCTTTGGAATTTATGACACAGTTGCTGTTTGGCAATATGA 480
Db 421 ATGATGACATTTGAACACCTTTGGAATTTATGACACAGTTGCTGTTTGGCAATATGA 480
Oy 481 CAATGTCAGTGAAGGCGAGAACTCTGCTGTGATGATGATGATGATGATGATGATG 540
Db 481 CAATGTCAGTGAAGGCGAGAACTCTGCTGTGATGATGATGATGATGATGATGATG 540
Oy 541 GGACCATAGTGTAAATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db 541 GGACCATAGTGTAAATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Oy 601 CTGTGGAAGTGAATATCCAGATGGAAGCGAATATCTGTCCATGGAATAGTTTG 660
Db 601 CTGTGGAAGTGAATATCCAGATGGAAGCGAATATCTGTCCATGGAATAGTTTG 660
Oy 661 GTGTCTCTCTACATGAGCAAAAGATTAATGATGAAGAGATGAGAGCAAAAGTGGATG 720
Db 661 GTGTCTCTCTACATGAGCAAAAGATTAATGATGAAGAGATGAGAGCAAAAGTGGATG 720
Oy 721 ACTGCCAGTTTGTCTGATAGCCCGACAGAAATTAAGTCTGATTTCTCAATCAAGTGA 780
Db 721 ACTGCCAGTTTGTCTGATAGCCCGACAGAAATTAAGTCTGATTTCTCAATCAAGTGA 780
Oy 781 AGAATCTGCAAAAGTTGAAGAGAGAGAGATTTTATGTTGAAGAGACCCGGAAC 840
Db 781 AGAATCTGCAAAAGTTGAAGAGAGAGAGATTTTATGTTGAAGAGACCCGGAAC 840
Oy 841 TTGATGACATGGAACAGCAAAAGGACATGTCATCAAGGTTACTCGAAGAGTTTA 900
Db 841 TTGATGACATGGAACAGCAAAAGGACATGTCATCAAGGTTACTCGAAGAGTTTA 900
Oy 901 CAATGATTTGGTGAAGAGCATTCAGTATGATGATCAACATTCATGAAGACTTCTGT 960
Db 901 CAATGATTTGGTGAAGAGCATTCAGTATGATGATCAACATTCATGAAGACTTCTGT 960
Oy 961 TGACCTATGAGACTTTTCTTTTCTAGCCCAATGGAATGGGCAAAAGATTATGAGATG 1020
Db 961 TGACCTATGAGACTTTTCTTTTCTAGCCCAATGGAATGGGCAAAAGATTATGAGATG 1020
Oy 1021 TTAATGACCCGAGCTCAGGAGTAAAGTTACACGGGTAGTATATGTTGGTAAATATC 1080
Db 1021 TTAATGACCCGAGCTCAGGAGTAAAGTTACACGGGTAGTATATGTTGGTAAATATC 1080
```

```
Db 1021 TTAAATGACCCGAGCCTCAGGGATAAAGTTACAAGGCTACTATTATTGGTGAATATC 1080
Qy 1081 ACTTCAATGACTTTGGAAGAGATCTGCAATGATCTGATTTTGTAGAGAATTTGAAACA 1140
Db 1081 ACTTCAATGACTTTGGAAGAGATCTGCAATGATCTGATTTTGTAGAGAATTTGAAACA 1140
Qy 1141 ATCTGGAAGAGAGAAATGGGTGACACCTTAAGGCTGTGTAATTCGGTGTGCTGCTA 1200
Db 1141 ATCTGGAAGAGAGAAATGGGTGACACCTTAAGGCTGTGTAATTCGGTGTGCTGCTA 1200
Qy 1201 AAGCAAAAGAGATGATGAGAGTTTAAACAACCATCCGAGAGTCCCTTGGCCCTTTA 1260
Db 1201 AAGCAAAAGAGATGATGAGAGTTTAAACAACCATCCGAGAGTCCCTTGGCCCTTTA 1260
Qy 1261 TCTTACTTGGAGGCTCTGAGAAAGGATTTGGAATCTTTGTGTGACAGTGTAGATTCAGTA 1320
Db 1261 TCTTACTTGGAGGCTCTGAGAAAGGATTTGGAATCTTTGTGTGACAGTGTAGATTCAGTA 1320
Qy 1321 GCAAAAGCACTGAAGCAGGCTTGAACGGGGGATCAGATATTAGAAGTAAATGGCCAA 1380
Db 1321 GCAAAAGCACTGAAGCAGGCTTGAACGGGGGATCAGATATTAGAAGTAAATGGCCAA 1380
Qy 1381 ACTTGAAGAACTTCAAGCTGTCAAAAGCTATGGAATTTCTTAGAATTAACACATTTAT 1440
Db 1381 ACTTGAAGAACTTCAAGCTGTCAAAAGCTATGGAATTTCTTAGAATTAACACATTTAT 1440
Qy 1441 CTATCACTGTGAACCAATTTATTTGTATTTAAAGACTTCTTACAGATTTCTCAGAG 1500
Db 1441 CTATCACTGTGAACCAATTTATTTGTATTTAAAGAACTTCTTACAGATTTCTCAGAG 1500
Qy 1501 AGAAAGAAATGGTGGCCCCCACCCTTCTAAATTTGGTGACATTTAAAGAGCCAGTGGCT 1560
Db 1501 AGAAAGAAATGGTGGCCCCCACCCTTCTAAATTTGGTGACATTTAAAGAGCCAGTGGCT 1560
Qy 1561 ACTCCATTCAGATCTTGTGTAGATGTAGAACAGTGTATGAGACTTGAAGAAAGTGAACA 1620
Db 1561 ACTCCATTCAGATCTTGTGTAGATGTAGAACAGTGTATGAGACTTGAAGAAAGTGAACA 1620
Qy 1621 AAAAAAGTAAAGCCACACGTGTGGAGGAAGAAACAAGCTGAAGAAATCTCCACAGAGA 1680
Db 1621 AAAAAAGTAAAGCCACACGTGTGGAGGAAGAAACAAGCTGAAGAAATCTCCACAGAGA 1680
Qy 1681 CTGGATCAGATATCTTGGCCACAGAAACCATATGATATTGGGATTTGGACGTCTCAG 1740
Db 1681 CTGGATCAGATATCTTGGCCACAGAAACCATATGATATTGGGATTTGGACGTCTCAG 1740
Qy 1741 ATGACAGCATAGTAGATTTAAGGACAGAAAGCAGATCCCAACTGCATTTGCCCTCAGTG 1800
Db 1741 ATGACAGCATAGTAGATTTAAGGACAGAAAGCAGATCCCAACTGCATTTGCCCTCAGTG 1800
Qy 1801 CAACCTTATCATCCAGTAAATCCGATTTTATGTCAGTCAATCATCGCATTTTATAGCTTCA 1860
Db 1801 CAACCTTATCATCCAGTAAATCCGATTTTATGTCAGTCAATCATCGCATTTTATAGCTTCA 1860
Qy 1861 GTGCTACTCCTGACCTGGCAGATCAAGTGTAAAGGTTTTTAAAGGCTGATACGAAAGCC 1920
Db 1861 GTGCTACTCCTGACCTGGCAGATCAAGTGTAAAGGTTTTTAAAGGCTGATACGAAAGCC 1920
Qy 1921 GCTACATCATGATGATAGAGACACTACAGCAAGAAAGTGTGATTCAGGCTATCAGGG 1980
Db 1921 GCTACATCATGATGATGATAGAGACACTACAGCAAGAAAGTGTGATTCAGGCTATCAGGG 1980
Qy 1981 AGTTTGTCTTACGTCACCCCGGATCAATATTCATATGTGAGAGTCTCTCTCACACCTG 2040
Db 1981 AGTTTGTCTTACGTCACCCCGGATCAATATTCATATGTGAGAGTCTCTCTCACACCTG 2040
Qy 2041 AGGGAATATCAACAAAGAGACTTCCAGATCAGCTTTCCAAACTTGCAGACAGAAATAC 2100
Db 2041 AGGGAATATCAACAAAGAGACTTCCAGATCAGCTTTCCAAACTTGCAGACAGAAATAC 2100
Qy 2101 AACTGAGTGAAGGATTTATCTGAAAAAACAACATGGAAGAACTCTTTGTTCCAGATG 2160
Db 2101 AACTGAGTGAAGGATTTATCTGAAAAAACAACATGGAAGAACTCTTTGTTCCAGATG 2160
```

```
Qy 2161 AAGATGCTCAGAGAGTGTGTGAGAGAGAGTCAATTTCCCTCCCTTCAGGTGAGCAGCTGGG 2220
Db 2161 AAGATGCTCAGAGAGTGTGTGAGAGAGAGTCAATTTCCCTCCCTTCAGGTGAGCAGCTGGG 2220
Qy 2221 AAGTTCGAACACAGCTCTCTATGCGAAATTTTGAACCTTTTCGACACATTTGAACCTACTG 2280
Db 2221 AAGTTCGAACACAGCTCTCTATGCGAAATTTTGAACCTTTTCGACACATTTGAACCTACTG 2280
Qy 2281 AATATATAGATGATTTATTTAACTCAGATCAAAAACACAGCTGTGCCAAGCTGAAAGAT 2340
Db 2281 AATATATAGATGATTTATTTAACTCAGATCAAAAACACAGCTGTGCCAAGCTGAAAGAT 2340
Qy 2341 TTGAAGAGTCTTTAACGAGAAACATTTTGGGTGACATCTGAAATTTCTCAGAGAAACAA 2400
Db 2341 TTGAAGAGTCTTTAACGAGAAACATTTTGGGTGACATCTGAAATTTCTCAGAGAAACAA 2400
Qy 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTTCATCAAGATAGCAGTGCAGTGTAGGG 2460
Db 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTTCATCAAGATAGCAGTGCAGTGTAGGG 2460
Qy 2461 AATGCAAGATTTTAACTCAATGTTTGCATATCATGATGAGTGGCTTAACCTGGCAGCAGTGG 2520
Db 2461 AATGCAAGATTTTAACTCAATGTTTGCATATCATGATGAGTGGCTTAACCTGGCAGCAGTGG 2520
Qy 2521 CAAGACTGGAGAGCAGCTGGAGAAACCTCCCATTAATACGAAACATATTTCAAGATC 2580
Db 2521 CAAGACTGGAGAGCAGCTGGAGAAACCTCCCATTAATACGAAACATATTTCAAGATC 2580
Qy 2581 TCCAGACCTGTTTGTATCTCCAGAAACATGCGAAATATGCTATGTTCTCAATAGTTC 2640
Db 2581 TCCAGACCTGTTTGTATCTCCAGAAACATGCGAAATATGCTATGTTCTCAATAGTTC 2640
Qy 2641 AAAATATCAACCTCCCATTAATCCCTCTATCCAGTTTCAAAAAGATCTCACCTTCC 2700
Db 2641 AAAATATCAACCTCCCATTAATCCCTCTATTTCCAGTTTCAAAAAGATCTCACCTTCC 2700
Qy 2701 TTCACGAGGAAATGACTCAAAAGTAGAGGAGTGTCAATTTTGAAGAGCTAAGAGATGA 2760
Db 2701 TTCACGAGGAAATGACTCAAAAGTAGAGGAGTGTCAATTTTGAAGAGCTAAGAGATGA 2760
Qy 2761 TTGCAAAAGAAATGCTCACGTGGCCGAATGAGCTTCAAGTGAACATGACCCCTGCCCTCA 2820
Db 2761 TTGCAAAAGAAATGCTCACGTGGCCGAATGAGCTTCAAGTGAACATGACCCCTGCCCTCA 2820
Qy 2821 TGTTCAGGACTGGAGAAAGAAATGGCGGAGTTTGGGCTCTCAGACCGAGGTAGTACAA 2880
Db 2821 TGTTCAGGACTGGAGAAAGAAATGGCGGAGTTTGGGCTCTCAGACCGAGGTAGTACAA 2880
Qy 2881 ATGCAACAGTGTAGATGTTTGTCTCAGACAGGTGTCAATAAAGCGGGTACGTCTAGTT 2940
Db 2881 ATGCAACAGTGTAGATGTTTGTCTCAGACAGGTGTCAATAAAGCGGGTACGTCTAGTT 2940
Qy 2941 CATTTCATTCGCCAAAAGCTTTATGAAGATCCCAATGGCTGCAAAAAGTGAAGCAAT 3000
Db 2941 CATTTCATTCGCCAAAAGCTTTATGAAGATCCCAATGGCTGCAAAAAGTGAAGCAAT 3000
Qy 3001 ACCTTTCAATTTGGAGCTGAGAAATGAGAGAGAGTCTTCAGACATTAATCTCTCAGT 3060
Db 3001 ACCTTTCAATTTGGAGCTGAGAAATGAGAGAGAGTCTTCAGACATTAATCTCTCAGT 3060
Qy 3061 GTGAGCCAGCAACACATTTGCTTGAAGATCTGTGACAAAAGAGCCGTGCAAAATCCG 3120
Db 3061 GTGAGCCAGCAACACATTTGCTTGAAGATCTGTGACAAAAGAGCCGTGCAAAATCCG 3120
Qy 3121 AGACCTTCCAGTAGTCTCCAAAGGCGAGGTCACAAAGAAAGTCTGCTCCGACAGC 3180
Db 3121 AGACCTTCCAGTAGTCTCCAAAGGCGAGGTCACAAAGAAAGTCTGCTCCGACAGC 3180
Qy 3181 CCCAGCAGCAGCCACACGACATTAATCAACGAGGACTTACAGGTTCCCGCGGTGT 3240
Db 3181 CCCAGCAGCAGCCACACGACATTAATCAACGAGGACTTACAGGTTCCCGCGGTGT 3240
```

OY 3241 CCCTTATCTTCACGGAAGAGTGCCTTAAGAGATCTCCACCTTTGGCATTAAC 3300
DB 3241 CCGTTATCTTCACGGAAGAGTGCCTTAAGAGATCTCCACCTTTGGCATTAAC 3300
OY 3301 CTCACAGCTTTAAAAAATTTCTTTGTCTGAAGAGAGAGTTTGAAGCTCACA 3360
DB 3301 CTCACAGCTTTAAAAAATTTCTTTGTCTGAAGAGAGAGTTTGAAGCTCACA 3360
OY 3361 AGAAGAGCTGAAGATCAATATCAATGCACTTGCAGCTTCTCTCTCTACTT 3420
DB 3361 AGAAGAGCTGAAGATCAATATCAATGCACTTGCAGCTTCTCTCTCTACTT 3420
OY 3421 CTCACAGAGTTCTTCCAGGAAGAGCTATCTTGGCTCCAGTGGTACTGTGATAAT 3480
DB 3421 CTCACAGAGTTCTTCCAGGAAGAGCTATCTTGGCTCCAGTGGTACTGTGATAAT 3480
OY 3481 TTTTCAGATTTCTGTACAGTGAATTTCTTACGATCCAGTAATTTAGCAATTCGTCTT 3540
DB 3481 TTTTCAGATTTCTGTACAGTGAATTTCTTACGATCCAGTAATTTAGCAATTCGTCTT 3540
OY 3541 TTTGACTGAGTGCAGTCTCAGTGCAGATGAGAGGCGCAGAGGCTTCTGTACAGATCG 3600
DB 3541 TTTGACTGAGTGCAGTCTCAGTGCAGATGAGAGGCGCAGAGGCTTCTGTACAGATCG 3600
OY 3601 TGGAAACAAACCTAGGAGTGGGACAGATGAGAGGCGGACATGATGAACTGATCAGT 3660
DB 3601 TGGAAACAAACCTAGGAGTGGGACAGATGAGAGGCGGACATGATGAACTGATCAGT 3660
OY 3661 ATAGCTTGGGGTCTTATGACCAATGTCGAGGCGGAGGCTTATATGCTAGAGCTACAG 3720
DB 3661 ATAGCTTGGGGTCTTATGACCAATGTCGAGGCGGAGGCTTATATGCTAGAGCTACAG 3720
OY 3721 TAAATTTCTTCCAGACAGAGAACTTTCCAGAGATCAGGAGGAGTGGCGGTCCACTTG 3780
DB 3721 TAAATTTCTTCCAGACAGAGAACTTTCCAGAGATCAGGAGGAGTGGCGGTCCACTTG 3780
OY 3781 ATGCTGTGACAGTGGCGGTGGGAGCTGAGCTCATGCTCAAGTGGCTCCATGATAATA 3840
DB 3781 ATGCTGTGACAGTGGCGGTGGGAGCTGAGCTCATGCTCAAGTGGCTCCATGATAATA 3840
OY 3841 TACAGACGATCCAGACACAGAGAACTGGGAGACTTTCATTCGGGATACCTCACTTGG 3900
DB 3841 TACAGACGATCCAGACACAGAGAACTGGGAGACTTTCATTCGGGATACCTCACTTGG 3900
OY 3901 ATTATTCAGGGGATCCTGAGGTTTATGGGACTCAAGCAGCATATGAGCCAAATTAATGT 3960
DB 3901 ATTATTCAGGGGATCCTGAGGTTTATGGGACTCAAGCAGCATATGAGCCAAATTAATGT 3960
OY 3961 TTTCTGATCATAGCAAAAGTATAACAGGCAAAATCAAAAGTAGAGAGGCTTGAACAAG 4020
DB 3961 TTTCTGATCATAGCAAAAGTATAACAGGCAAAATCAAAAGTAGAGAGGCTTGAACAAG 4020
OY 4021 CCCATTCGCCAGACAGCTGGGCGTCTTCCACAGGTTACTGGGGAACAAGCTAGAAAGTGT 4080
DB 4021 CCCATTCGCCAGACAGCTGGGCGTCTTCCACAGGTTACTGGGGAACAAGCTAGAAAGTGT 4080
OY 4081 ACACAGGCAATTAAGGAGGAGGAGGAGGAGGATGTTTCCATTGAAGCCGAAGACAGTA 4140
DB 4081 ACACAGGCAATTAAGGAGGAGGAGGAGGAGGAGGATGTTTCCATTGAAGCCGAAGACAGTA 4140
OY 4141 GCCTAAGCTGTGTGACTAGGAAGAACCAAGCTGTCCCATGCTGTGCCCATAGCTGTG 4200
DB 4141 GCCTAAGCTGTGTGACTAGGAAGAACCAAGCTGTCCCATGCTGTGCCCATAGCTGTG 4200
OY 4201 TGGCATCAAGTACTACAAAGGGGCTCATTTGCAAGAAAGAGGAGGATGCAAGAGCCCC 4260
DB 4201 TGGCATCAAGTACTACAAAGGGGCTCATTTGCAAGAAAGAGGAGGATGCAAGAGCCCC 4260
OY 4261 CGCCACCCCTTCGGGCTACATTTGAATTCATTTGAGTCTTCCAGAAAGGAGGACTGCC 4320
DB 4261 CGCCACCCCTTCGGGCTACATTTGAATTTGAGTCTTCCAGAAAGGAGGACTGCC 4320
OY 4321 ATTCAGCCAGGAAGACGCGGAGCTACAAAGGTGGCCCTTCAGAGATGGGAGTGTGCAC 4380

DB 4321 ATTCAGCCAGGAACCGCGGAGCTACAAAGTGGCCCTTCAGAGATGGGAGTGTGCAC 4380
OY 4381 GATCTCCGACACACACTGGGCTTATCGGTACAGACAGCCACATGGGCAATCCACACAGA 4440
DB 4381 GATCTCCGACACACACTGGGCTTATCGGTACAGACAGCCACATGGGCAATCCACACAGA 4440
OY 4441 GCAGGCTGTGAAACAACCTCAGTGGCATTAACCGAAGAGCTGACCGGCGCTGGCCC 4500
DB 4441 GCAGGCTGTGAAACAACCTCAGTGGCATTAACCGAAGAGCTGACCGGCGCTGGCCC 4500
OY 4501 CTATACGTCCTCCAGGAGTCTTCCACCGAGAGAGATGAAGATGAAGTTCGTGTTT 4560
DB 4501 CTATACGTCCTCCAGGAGTCTTCCACCGAGAGAGATGAAGATGAAGTTCGTGTTT 4560
OY 4561 GAGGACACAGACTTTCTTGGAGACAGAGGAGCGCCACTGAAAGAGAGACAAAGAGAGCT 4620
DB 4561 GAGGACACAGACTTTCTTGGAGACAGAGGAGCGCCACTGAAAGAGAGACAAAGAGAGCT 4620
OY 4621 CCTGAGCATTTGGAGGCTTGGAACTCACAATTCAGAGAGGAGTGGACCAAGTTCCTCTTC 4680
DB 4621 CCTGAGCATTTGGAGGCTTGGAACTCACAATTCAGAGAGGAGTGGACCAAGTTCCTCTTC 4680
OY 4681 CCTGCTTAAAGACAGCATGGGGCTTCTTCCCTTCCCTTCCCTTGCATGTGA 4740
DB 4681 CCTGCTTAAAGACAGCATGGGGCTTCTTCCCTTCCCTTCCCTTGCATGTGA 4740
OY 4741 AATACTGTGAAGAAATTTGCCCTGGCAGCTTTTACAGACTTTGTTGCTTGAATGCACAGTGC 4800
DB 4741 AATACTGTGAAGAAATTTGCCCTGGCAGCTTTTACAGACTTTGTTGCTTGAATGCACAGTGC 4800
OY 4801 AGCAATCTTCGAGCTCCACAGTGTGCTGCTGCACATACACAGATATTCGCAATTC 4860
DB 4801 AGCAATCTTCGAGCTCCACAGTGTGCTGCTGCACATACACAGATATTCGCAATTC 4860
OY 4861 CAAGATCATCACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4920
DB 4861 CAAGATCATCACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4920
OY 4921 TTTTAAATCTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 4980
DB 4921 TTTTAAATCTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 4980
OY 4981 AGAAAGCTAGGCAATGAACTCTTGGGGCTTTAAACCCAGGAAGACAAAGAAAA 5040
DB 4981 AGAAAGCTAGGCAATGAACTCTTGGGGCTTTAAACCCAGGAAGACAAAGAAAA 5040
OY 5041 CAATGAATCCTTTGAGTACAGTGTGTCTCACTTGTTCATGATGCTCTTTTAAAAA 5100
DB 5041 CAATGAATCCTTTGAGTACAGTGTGTCTCACTTGTTCATGATGCTCTTTTAAAAA 5100
OY 5101 AAAAAATGAGTTTAAAGATTTTGTTCAGAGAGTAAATATATATTCATTTAAAGATTACAG 5160
DB 5101 AAAAAATGAGTTTAAAGATTTTGTTCAGAGAGTAAATATATATTCATTTAAAGATTACAG 5160
OY 5161 TATTAATTTTAAACCTTAAGTAGGTTGCCAGGCTGTTTCTGAAAAAACCAAAATATGCGCG 5220
DB 5161 TATTAATTTTAAACCTTAAGTAGGTTGCCAGGCTGTTTCTGAAAAAACCAAAATATGCGCG 5220
OY 5221 ACAGGGTGTGGCCACACCAAGAAAGCGGGAAGAGCTGTGACCTTGCCCTCATG 5280
DB 5221 ACAGGGTGTGGCCACACCAAGAAAGCGGGAAGAGCTGTGACCTTGCCCTCATG 5280
OY 5281 TCCCTTCTGGTCTCACCGGAGAGTGCCTTATCTCTGGAAGTAAATGTTTACCAATTA 5340
DB 5281 TCCCTTCTGGTCTCACCGGAGAGTGCCTTATCTCTGGAAGTAAATGTTTACCAATTA 5340
OY 5341 TACCAAGACACCTCATTTGCTCTTCCAGAGTGGGCTTCTGTGAAAACTGTTTG 5400
DB 5341 TACCAAGACACCTCATTTGCTCTTCCAGAGTGGGCTTCTGTGAAAACTGTTTG 5400
OY 5401 CACATGGCCAGGAGGAGGAATAGAGCCCTTGTCTCTGTGAGGCTTATGAGGACAG 5460

Db 5401 CACATGGCCAGGGGAGGAAGAACTAGACACCTGTGTCTGTCTGAGCCTTATGAGGACG 5460
Qy 5461 ACAGTGTCAATGGCGGATGTGCTTCCTCATTTGAGATGGATGGCAACCCCATTTTAA 5520
Db 5461 ACAGTGTCAATGGCGGATGTGCTTCCTCATTTGAGATGGATGGCAACCCCATTTTAA 5520
Qy 5521 GTATATATTTCTTTGATTTTGTAAATTTAGAGGTGATGTTTGTGTTTGTGTTTGTGTT 5580
Db 5521 GTATATATTTCTTTGATTTTGTAAATTTAGAGGTGATGTTTGTGTTTGTGTTTGTGTTT 5580
Qy 5581 TTTTAAAGAGAAACATTTATTAACAGGATGAGTGAACACAGCTTGGATGTT 5640
Db 5581 TTTTAAAGAGAAACATTTATTAACAGGATGAGTGAACACAGCTTGGATGTT 5640
Qy 5641 GGAGCTAATGCCAGCTGTTTATCTGCTCTTTCAGACAGCCTCCCTTATTTGAATTGGC 5700
Db 5641 GGAGCTAATGCCAGCTGTTTATCTGCTCTTTCAGACAGCCTCCCTTATTTGAATTGGC 5700
Qy 5701 ATTAGGGAATTAACAAGCCTTAAAGCTGATTAAGATCAAAAACCTGGTGAACATGGC 5760
Db 5701 ATTAGGGAATTAACAAGCCTTAAAGCTGATTAAGATCAAAAACCTGGTGAACATGGC 5760
Qy 5761 AGCCTTGAAGGAGGATTAAGTACCAAAAGCTAACCTCCAGTGGCTTATGAGAGCTG 5820
Db 5761 AGCCTTGAAGGAGGATTAAGTACCAAAAGCTAACCTCCAGTGGCTTATGAGAGCTG 5820
Qy 5821 CATATAGAGAAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGCTATTAACCTTAAAT 5880
Db 5821 CATATAGAGAAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGCTATTAACCTTAAAT 5880
Qy 5881 GACGGAATAACCCCTCCACTGATTTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 5940
Db 5881 GACGGAATAACCCCTCCACTGATTTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 5940
Qy 5941 AGCCTGCCAATCTGAGTACTCAATCTGAGAACTGCTGTGCTTGTGAATTTTTC 6000
Db 5941 AGCCTGCCAATCTGAGTACTCAATCTGAGAACTGCTGTGCTTGTGAATTTTTC 6000
Qy 6001 ATTAATATTCAGCTGATCATATGATGATGATGATGATGATGATGATGATGATGATGATG 6060
Db 6001 ATTAATATTCAGCTGATCATATGATGATGATGATGATGATGATGATGATGATGATGATG 6060
Qy 6061 GTGGAATTCAGCTGTTTTCACGTATCAAAACATGTCAGTGGCTTATTTAATTTCT 6120
Db 6061 GTGGAATTCAGCTGTTTTCACGTATCAAAACATGTCAGTGGCTTATTTAATTTCT 6120
Qy 6121 CTCTGTATCATGSCATTTGTCTACTGCTTATTAATGATGATGATGATGATGATGATGATG 6180
Db 6121 CTCTGTATCATGSCATTTGTCTACTGCTTATTAATGATGATGATGATGATGATGATGATG 6180
Qy 6181 TTACATGTAATATGATTTATTTGTCAGTTTATTAATGAGCTATGAGACCTCATGTGAT 6240
Db 6181 TTACATGTAATATGATTTATTTGTCAGTTTATTAATGAGCTATGAGACCTCATGTGAT 6240
Qy 6241 ATAGAAGACAGAAATCTAGCTACGACAAAGTGCACAAATGTTATCTAAGCAATTAAGT 6300
Db 6241 ATAGAAGACAGAAATCTAGCTACGACAAAGTGCACAAATGTTATCTAAGCAATTAAGT 6300
Qy 6301 AATGTAGACATAGAGCTGCTAATCTCAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360
Db 6301 AATGTAGACATAGAGCTGCTAATCTCAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360
Qy 6361 CAATTAATGCTGATTCCTCATACCTTTGATGATGATGATGATGATGATGATGATGATGATG 6420
Db 6361 CAATTAATGCTGATTCCTCATACCTTTGATGATGATGATGATGATGATGATGATGATGATG 6420
Qy 6421 AGACATTTGGTGGAGTCTGATCCCTTTGATTTTAAATTAATTAATTAATTAATTAATTAAT 6480
Db 6421 AGACATTTGGTGGAGTCTGATCCCTTTGATTTTAAATTAATTAATTAATTAATTAATTAAT 6480
Qy 6481 TATATTTTGTGGAAGATGATGAATGATGATGATGATGATGATGATGATGATGATGATGATG 6540
Db 6481 TATATTTTGTGGAAGATGATGAATGATGATGATGATGATGATGATGATGATGATGATGATG 6540

Qy 6541 CACGCTGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6568
Db 6541 CACGCTGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6568

RESULT 3
AB002311 6568 bp mRNA linear PRI 06-OCT-2001
LOCUS Human mRNA for KIAA0313 gene, complete cds.
DEFINITION AB002311
ACCESSION AB002311.1 GI:2224566
VERSION AB002311.1
KEYWORDS Homo sapiens male brain cDNA to mRNA, clone.lib:pbuescriptII SK plus clone:H0186.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O. Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro
DNA Res. 4 (2), 141-150 (1997)
97349984
2 (bases 1 to 6568)
Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
Direct Submission
Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@fokazusa.or.jp, Tel:+81-438-52-3913)
Location/Qualifiers

FEATURES
source
1..6568
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="H0186"
/sex="male"
/tissue-type="brain"
/clone.lib="pbuescriptII SK plus"
1..6568
/gene="KIAA0313"
63..4562
/gene="KIAA0313"
/codon.start=1
/protein_id="BAA20772.1"
/db_xref="GI:2224567"
/translation="MKPLAIPANHGVMGQGRKSLPADFTKLHLTLDLSLHPQVTHVSSS HSGCITSSGSSLSIDYQATESPAGMDLSGLPETAVIDSEDDDEIERADPLM SMDIVDCLDKPDIRDDDIETOLEPMLQIPAFANMTMSVRELCAVMFAVVERAG TIVLDGEILDWSVILNGSVEVITYPDKAEILLCNGNSFGVSP7MDKRYTKGVARTY DDCQFVIAQOQYCRILNOEKRMQKVEEBEIVWVKRHLRDLRTGRKHIVLRGTS ERLTHLVEHSEVDPTEIEDELIVRFPLSPMEVGKLLLEWNPDSLRDKVTVVYL LVMNHPNDFEGDPAMTRFLFEENLRREKMGHRLRLNIAKAKARLRLMTLRPS REAPLPFILGSEKGFGLFVDSVDSGSKATGKRGDOLLEVANGQFENIOLSKAM ELRLNTHLSTVKTNLVFEKELRLSEBRKNAPHLPKIGDIKKARSYSLPDLAV VOQVIGLEKVNKSKRANTVGGKRLKLLDTRISILQKRPNDIGIOSDDSLVGL RQTKHLPALPVSGTSSNBDLSQSHRITLDFSATPDLPOVLVEFVADOOSRYTMI SKDTAKEVYIOAIRFATVATPDQSLCEVSPVEGYIKQRLRPDQLSLADRIQLS GRVYLKNNMETTLCSDDEADQLRESQSLSLSTVAVATOLSKRNFELRNIEPTE YTDDELFLKLSKTSKNSCANIKREFEVIQETFMVASELRENNOLKRMKLIKFKIALHC RECKNFSMFALISGILNAPVARLITMEKLPNKYEKLPQDQDLFDSRMKARVY LNSQNLPPILFLPVIKDLTFLHEGNDKVDGLVNEKRLMKILKEIRHYGRASVY MPALMFRTRKKRWSLSLSQSTNATVLDVAQGTGKRRKRVRRSSFLNARKLVDAQ MARKVOYLSNLELDEESLQTLSDCEPATNLPKNKCDKPKVKSFTSVAPARVY IAPMSGRGLVATVATVISPSTEELSDQGDASLDADSGKSTSCSSSHNDIOT IOHORSWETLPGHTHFDYSDGPAGLMSSSMDQIMKSDHSTKYRQNSRESLEQA OSRAWSASTGYWGEDESDGTIKRGGKDVSIIESSSLTSCVTEETKPKVPAPAI AVASSTTKGLIARKEGRYREPPGTGGYIGITDPEGSHAPARKPPDYVVALQRSR MARSDTAGPSSVQPHGHPITSRNPKNPKHKNESDRLAPVQSGFSTEDEDE

Db 2041 AGGAGTAACTCAAAACAAGACCTTCAGATTCAGCTTTCACAACTTGGACAGAGATAC 2100
Oy 2101 AACTGAGTGGAGAGTATATATCTGAAAAACACATGSAACAGAACTCTTTGTTGATG 2160
Db 2101 AACTGAGTGGAGAGTATATATCTGAAAAACACATGSAACAGAACTCTTTGTTGATG 2160
Oy 2161 AGATGCTCAGAGGTTGTGAGAGAGATCAATTTCCCTCCCTTCAGCTCAGACAGTGG 2220
Db 2161 AGATGCTCAGAGGTTGTGAGAGAGATCAATTTCCCTCCCTTCAGCTCAGACAGTGG 2220
Oy 2221 AAGTTGCAACACAGCTCTCTATGCGAAATTTGAACTCTTTCGCAACATTTGAACCTACTG 2280
Db 2221 AAGTTGCAACACAGCTCTCTATGCGAAATTTGAACTCTTTCGCAACATTTGAACCTACTG 2280
Oy 2281 AATATATAGATGATTTATTTTAACTCAGATCAAAAACACAGCTCTGCCAACCTTAAGAT 2340
Db 2281 AATATATAGATGATTTATTTTAACTCAGATCAAAAACACAGCTCTGCCAACCTTAAGAT 2340
Oy 2341 TTGAGAGATCAATTAACCCAGAAACATTTGGGAGACATGGAATTTCTCAGAGAAACA 2400
Db 2341 TTGAGAGATCAATTAACCCAGAAACATTTGGGAGACATGGAATTTCTCAGAGAAACA 2400
Oy 2401 ACCAGCTGAGAGAGATGAGATCATTAAGCATTTTCATCAAGATAGCACTGCATGTAGGG 2460
Db 2401 ACCAGCTGAGAGAGATGAGATCATTAAGCATTTTCATCAAGATAGCACTGCATGTAGGG 2460
Oy 2461 AATGCAAGATTTTAACTCAATGTTTGCATCACTGACCTGACACCTGGACCCAGTGG 2520
Db 2461 AATGCAAGATTTTAACTCAATGTTTGCATCACTGACCTGACACCTGGACCCAGTGG 2520
Oy 2521 CAAGACTGCGAAGCACTGGAGAACTTCCCAATAAATACGAAAACTATTTTCAGATC 2580
Db 2521 CAAGACTGCGAAGCACTGGAGAACTTCCCAATAAATACGAAAACTATTTTCAGATC 2580
Oy 2581 TCCAAAGACCTGTTTATCTTCCAGAAACATGSCAAATATCTGATGTTCTCAATAGTC 2640
Db 2581 TCCAAAGACCTGTTTATCTTCCAGAAACATGSCAAATATCTGATGTTCTCAATAGTC 2640
Oy 2641 AAAATCTCAACCTCCCAATAATCCCTCTATTTCCAGTTATCAAAAAGATCTCCCTCC 2700
Db 2641 AAAATCTCAACCTCCCAATAATCCCTCTATTTCCAGTTATCAAAAAGATCTCCCTCC 2700
Oy 2701 TTCAAGAGAAATGACATCAAAATAGAGGGCTGTCATTTTGAAGCTTAAGATGA 2760
Db 2701 TTCAAGAGAAATGACATCAAAATAGAGGGCTGTCATTTTGAAGCTTAAGATGA 2760
Oy 2761 TTGCAAAAAGAAATTCGTACGTTGGCCGATGCTTCAAGTACATGACACCTCCCTCA 2820
Db 2761 TTGCAAAAAGAAATTCGTACGTTGGCCGATGCTTCAAGTACATGACACCTCCCTCA 2820
Oy 2821 TGTTCAGAGACTCGAAGAAATGGCGGAGTTTGGGGTCTCTAGCCAGGGTAGTACAA 2880
Db 2821 TGTTCAGAGACTCGAAGAAATGGCGGAGTTTGGGGTCTCTAGCCAGGGTAGTACAA 2880
Oy 2881 ATGCAACAGTCTAGATGTGCTCAGACAGGTGTCATTAATAAGCGGGTAGTGTAGT 2940
Db 2881 ATGCAACAGTCTAGATGTGCTCAGACAGGTGTCATTAATAAGCGGGTAGTGTAGT 2940
Oy 2941 CCTTTCTCAATGCGAAAAAGCTTTATGAAGATGCGCAATGGCTCGAAAAATGGAAGCAGT 3000
Db 2941 CCTTTCTCAATGCGAAAAAGCTTTATGAAGATGCGCAATGGCTCGAAAAATGGAAGCAGT 3000
Oy 3001 ACCTTTCAATTTGGAGTAGAATGAGAGAGAGTCTTCAGACATTTATCTCTGCAGT 3060
Db 3001 ACCTTTCAATTTGGAGTAGAATGAGAGAGAGTCTTCAGACATTTATCTCTGCAGT 3060
Oy 3061 GTGAGCCAGCAACAACATTTGCTTAAGAAATCTGTGACAAAAAGCGCTGTCAAAATCG 3120
Db 3061 GTGAGCCAGCAACAACATTTGCTTAAGAAATCTGTGACAAAAAGCGCTGTCAAAATCG 3120
Oy 3121 AGACCTTCTCAGTACTCCAGAGGAGGCTCACAACAAAAAGCTCAGTCCCTGCACAGC 3180
Db 3121 AGACCTTCTCAGTACTCCAGAGGAGGCTCACAACAAAAAGCTCAGTCCCTGCACAGC 3180
Db 3121 AGACCTTCTCAGTACTCCAGAGGAGGCTCACAACAAAAAGCTCAGTCCCTGCACAGC 3180

Oy 3181 CCCAGAGAGCCACACACACATTAATAATCAACCCAGGACTACAGGTTCCCGCGTGT 3240
Db 3181 CCCAGAGAGCCACACACACATTAATAATCAACCCAGGACTACAGGTTCCCGCGTGT 3240
Oy 3241 CCCATTATCTTACAGGAGAAAGTCCGTAAGAGATCTCCACCTTTTGGCATAACT 3300
Db 3241 CCCATTATCTTACAGGAGAAAGTCCGTAAGAGATCTCCACCTTTTGGCATAACT 3300
Oy 3301 CTCACACAGCTTTAAAAAATTTCTTCTGCTGGAAGAGAAAGTTTGGAGCTGCA 3360
Db 3301 CTCACACAGCTTTAAAAAATTTCTTCTGCTGGAAGAGAAAGTTTGGAGCTGCA 3360
Oy 3361 AGAAACAGGCTGAAGATCAATATCAATATGCAATCTTCCAGCTTTCTTCTCTACTT 3420
Db 3361 AGAAACAGGCTGAAGATCAATATCAATATGCAATCTTCCAGCTTTCTTCTCTACTT 3420
Oy 3421 CTCACAGAGTTCTCCAGAGAAAGCTATCTTTGGCTCCAGTGTACTGTGATAT 3480
Db 3421 CTCACAGAGTTCTCCAGAGAAAGCTATCTTTGGCTCCAGTGTACTGTGATAT 3480
Oy 3481 TTTCAAGATCTGCTCAGTCAAGTAAATTTCTTCAAGATCTTCAAGATCTTCAAGAT 3540
Db 3481 TTTCAAGATCTGCTCAGTCAAGTAAATTTCTTCAAGATCTTCAAGATCTTCAAGAT 3540
Oy 3541 TTGACTCAGTGGCAGTCTCAGTCAAGATGAGAGGCGCCAGAGGATCTGTCAAGATC 3600
Db 3541 TTGACTCAGTGGCAGTCTCAGTCAAGATGAGAGGCGCCAGAGGATCTGTCAAGATC 3600
Oy 3601 TGGAAACAAACCTAGGAGATGGGAGATGAGAGGCGGCAATGATGAACTGATAGT 3660
Db 3601 TGGAAACAAACCTAGGAGATGGGAGATGAGAGGCGGCAATGATGAACTGATAGT 3660
Oy 3661 ATGAGTGGGGTCTATGACCAATGTCCGAGGCGGAGGCTTATATGCTTACAGCTACG 3720
Db 3661 ATGAGTGGGGTCTATGACCAATGTCCGAGGCGGAGGCTTATATGCTTACAGCTACG 3720
Oy 3721 TAAATTTCTTCCACAGAGAGAACTTCCAGAGATGAGGGGATCGCGCTCACTTG 3780
Db 3721 TAAATTTCTTCCACAGAGAGAACTTCCAGAGATGAGGGGATCGCGCTCACTTG 3780
Oy 3781 ATGCTGCTCAGATGCGCGGTGGAGCTGAGAGCTGATGCTCAAGTGGCTCCATGATAT 3840
Db 3781 ATGCTGCTCAGATGCGCGGTGGAGCTGAGAGCTGATGCTCAAGTGGCTCCATGATAT 3840
Oy 3841 TACAGAGATCCAGACACAGAGAGCTGGAGACTCTTCATTCGGGCACTACTCTTG 3900
Db 3841 TACAGAGATCCAGACACAGAGAGCTGGAGACTCTTCATTCGGGCACTACTCTTG 3900
Oy 3901 ATTATTCAGGGATCTGCAAGTTTATGGGATCAAGCAGCTATATGACCAATATATG 3960
Db 3901 ATTATTCAGGGATCTGCAAGTTTATGGGATCAAGCAGCTATATGACCAATATATG 3960
Oy 3961 TTTTCTGATCTACACCAAAATATACAGGCAAAATCAAAATGATAGAGAGAGCTTGAACAAG 4020
Db 3961 TTTTCTGATCTACACCAAAATATACAGGCAAAATCAAAATGATAGAGAGAGCTTGAACAAG 4020
Oy 4021 CCCAGTCCGAGAAAGCTGGCGCTTCTCCACAGATTACTGGGAGAAAGCTCAGAAAGTG 4080
Db 4021 CCCAGTCCGAGAAAGCTGGCGCTTCTCCACAGATTACTGGGAGAAAGCTCAGAAAGTG 4080
Oy 4081 ACACAGGACATTAAGCGGAGGAGGAGATGTTTCAATTGAAGCGGAAGAGTA 4140
Db 4081 ACACAGGACATTAAGCGGAGGAGGAGATGTTTCAATTGAAGCGGAAGAGTA 4140
Oy 4141 GCTTAAGCTGTGATACGAGAAACCAAGGCTGTCCCATGCTCTCCACATAGCTG 4200
Db 4141 GCTTAAGCTGTGATACGAGAAACCAAGGCTGTCCCATGCTCTCCACATAGCTG 4200
Oy 4201 TGGCATCAAGTACTCAAAAGGGGCTATTGTCAGAAAGGAGGAGATTCAGAGAGCCC 4260
Db 4201 TGGCATCAAGTACTCAAAAGGGGCTATTGTCAGAAAGGAGGAGATTCAGAGAGCCC 4260

4261 CGCCACCCCTCCGGCTACATTGGAATTCCTTACCTGACTTTCAGAGAAGGCACTGCC 4320
4261 CGCCACCCCTCCGGCTACATTGGAATTCCTTACCTGACTTTCAGAGAAGGCACTGCC 4320
4321 ATCCACCCAGGAAGCCGGGCACTACACGCTGGCCCTTGAGAGATGCGGATGCTGCAC 4380
4321 ATCCACCCAGGAAGCCGGGCACTACACGCTGGCCCTTGAGAGATGCGGATGCTGCAC 4380
4381 GATCCCTCCAGACAGCTGGGGCTTCATCCGCTACAGAGCCACATGGGCAATCCACACAGA 4440
4381 GATCCCTCCAGACAGCTGGGGCTTCATCCGCTACAGAGCCACATGGGCAATCCACACAGA 4440
4441 GCAGGCTGTGAACAACCTCAGTGGCATTAACCCAGAGCTGACCCGCGCTCGGCC 4500
4441 GCAGGCTGTGAACAACCTCAGTGGCATTAACCCAGAGCTGACCCGCGCTCGGCC 4500
4501 CTTATCAGTCCCAAGGGTTTCCACCGAGAGAGATGAAGATGAACAAGTTTGTGCTGTTT 4560
4501 CTTATCAGTCCCAAGGGTTTCCACCGAGAGAGATGAAGATGAACAAGTTTGTGCTGTTT 4560
4561 GAGGCACACACTTTTCTGAGACAGACGACCTGAAAGAGAGACACAGAGAGCT 4620
4561 GAGGCACACACTTTTCTGAGACAGACGACCTGAAAGAGAGACACAGAGAGCT 4620
4621 CGGACCATTTGGAGCCTTGGAACTCATTCTGAGAGAGGTGGACAGTTTGGCTCTTC 4680
4621 CGGACCATTTGGAGCCTTGGAACTCATTCTGAGAGAGGTGGACAGTTTGGCTCTTC 4680
4681 CCTGCCTTAAAGACACATGGGGCTTCTCCCTCTTCTTCCCTTCCCTTTCATGATGA 4740
4681 CCTGCCTTAAAGACACATGGGGCTTCTCCCTCTTCTTCCCTTTCATGATGA 4740
4741 AATACGTGAAGAATTTGCCCTGCGCACTTTTACAGACTTTGTTGCTTGAATGCAAGTGC 4800
4741 AATACGTGAAGAATTTGCCCTGCGCACTTTTACAGACTTTGTTGCTTGAATGCAAGTGC 4800
4801 AGCAATCTTGAAGCTCCCACTGTTGCTGCTGCGCATCAGACAGATATTCCTCAATTC 4860
4801 AGCAATCTTGAAGCTCCCACTGTTGCTGCTGCGCATCAGACAGATATTCCTCAATTC 4860
4861 CAAATCTTGAAGCTCCCACTGTTGCTGCTGCGCATCAGACAGATATTCCTCAATTC 4920
4861 CAAATCTTGAAGCTCCCACTGTTGCTGCTGCGCATCAGACAGATATTCCTCAATTC 4920
4921 TTTTATATCTCTCTTAAATTTCAATTCAGTCCAGTCCAGCTTGAATGCTGATGGAATG 4980
4921 TTTTATATCTCTCTTAAATTTCAATTCAGTCCAGTCCAGCTTGAATGCTGATGGAATG 4980
4981 AGAAAGCTAGCATTTGAAGTACTTGGGGCTTTAACCACCAAGGAAGACAAAGAAAA 5040
4981 AGAAAGCTAGCATTTGAAGTACTTGGGGCTTTAACCACCAAGGAAGACAAAGAAAA 5040
4981 AGAAAGCTAGCATTTGAAGTACTTGGGGCTTTAACCACCAAGGAAGACAAAGAAAA 5040
5041 CAATGAATCTTGAAGTACTTGGGGCTTTAACCACCAAGGAAGACAAAGAAAA 5100
5041 CAATGAATCTTGAAGTACTTGGGGCTTTAACCACCAAGGAAGACAAAGAAAA 5100
5101 AAAAATGAGTTTAAAGATTTGTTGAGAGAGTAAATATATATATCCATTTAATGATTAAG 5160
5101 AAAAATGAGTTTAAAGATTTGTTGAGAGAGTAAATATATATATCCATTTAATGATTAAG 5160
5161 TTTTATTTTAAACCTTAAGTGGGTGCGAGCCTGTTCTGAAAAACCAAAATATGCGCG 5220
5161 TTTTATTTTAAACCTTAAGTGGGTGCGAGCCTGTTCTGAAAAACCAAAATATGCGCG 5220
5221 ACAGGCTGGGCGACACAG 5280
5221 ACAGGCTGGGCGACACAG 5280
5281 TCGTTTGTGCTACCGCGGAGAGTGGCTTATCTGGAAGATGAATGTTAGCAATTTAA 5340
5281 TCGTTTGTGCTACCGCGGAGAGTGGCTTATCTGGAAGATGAATGTTAGCAATTTAA 5340
5341 TACGAAGACACCTCATCTCTCTTCCAGAGTGGATGGGGTCTTGTGAAAACTGTTTG 5400

5341 TACGAAGACACCTCATCTCTCTTCCAGAGTGGATGGGGTCTTGTGAAAACTGTTTG 5400
5401 CACATGGCCAGGGGAGGAGTACAGACCTGTGTCTGTCTGTAGAGCTTATGAGAGCAG 5460
5401 CACATGGCCAGGGGAGGAGTACAGACCTGTGTCTGTCTGTAGAGCTTATGAGAGCAG 5460
5401 CACATGGCCAGGGGAGGAGTACAGACCTGTGTCTGTCTGTAGAGCTTATGAGAGCAG 5460
5461 ACAGTGTCAATGGGAGATGTGCTCTCTCATTTGAATGAGTGGCAAAACCCCAATTTTAA 5520
5461 ACAGTGTCAATGGGAGATGTGCTCTCTCATTTGAATGAGTGGCAAAACCCCAATTTTAA 5520
5521 GTTATATTTCTTGAATTTTGTAAATTTAGAGTGGAGTTTGTGTTTGTGTTTGTGTT 5580
5521 GTTATATTTCTTGAATTTTGTAAATTTAGAGTGGAGTTTGTGTTTGTGTTTGTGTT 5580
5581 TTTTATTAAGAGAAACATTTATATAGTGAATGAGATTCGAGTGAAGCAGCTTGGATGTT 5640
5581 TTTTATTAAGAGAAACATTTATATAGTGAATGAGATTCGAGTGAAGCAGCTTGGATGTT 5640
5641 GAGCTAATGCCAGCTTTTATATCTGCTCTTTCAGACAGCTCCCTTTATGAAATGGC 5700
5641 GAGCTAATGCCAGCTTTTATATCTGCTCTTTCAGACAGCTCCCTTTATGAAATGGC 5700
5701 ATTAGGGAATTAACAAGCCTTTAAAGTGAATGAATGAATGAATGAATGAATGAATGAAT 5760
5701 ATTAGGGAATTAACAAGCCTTTAAAGTGAATGAATGAATGAATGAATGAATGAATGAAT 5760
5761 AGCCTTTGCAAGGACAGTGTAGTACCAAGACATCAACCTCAAGTGGCTTATGAGAGCTG 5820
5761 AGCCTTTGCAAGGACAGTGTAGTACCAAGACATCAACCTCAAGTGGCTTATGAGAGCTG 5820
5821 CATATGAGAGAGCTTAAAGTGTAGACACATCTGCTACAGCTGCTATTAACCTTAAT 5880
5821 CATATGAGAGAGCTTAAAGTGTAGACACATCTGCTACAGCTGCTATTAACCTTAAT 5880
5881 GACTGAATGAGAGCTTAAAGTGTAGACACATCTGCTACAGCTGCTATTAACCTTAAT 5940
5881 GACTGAATGAGAGCTTAAAGTGTAGACACATCTGCTACAGCTGCTATTAACCTTAAT 5940
5941 AGGCTCCCAATCTGAGTACTGATCAATGAGAGAACTGCTGCTGATTTTGTTC 6000
5941 AGGCTCCCAATCTGAGTACTGATCAATGAGAGAACTGCTGCTGATTTTGTTC 6000
6001 ATTAAATGAGTGTATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 6060
6001 ATTAAATGAGTGTATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 6060
6061 GTTGAATGAGTGTATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 6120
6061 GTTGAATGAGTGTATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 6120
6121 CTTCTGATCATGAGCTTGTGCTACTGCTTATTAATGATGATGATGATGATGATGATGAT 6180
6121 CTTCTGATCATGAGCTTGTGCTACTGCTTATTAATGATGATGATGATGATGATGATGAT 6180
6181 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6240
6181 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6240
6241 ATGAAGAGAGAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6300
6241 ATGAAGAGAGAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6300
6301 AATTGTAGAACATGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6360
6301 AATTGTAGAACATGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6360
6361 CAATTAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6420
6361 CAATTAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6420
6421 AGCATTTGAGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6480
6421 AGCATTTGAGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6480

misc_feature 12156. .12567
/note="match to EST BG206034 (NID:g137272721)"
misc_feature 12157. .12567
/note="match to EST BG199786 (NID:g13721473)"
misc_feature 12366. .12543
/note="match to EST AW467230 (NID:g7037336) he07g06.x1"
misc_feature 12366. .12499
/note="similar to Homo sapiens EST N58568 (NID:g1202458)
yvs5c09.r1"
misc_feature 12367. .12539
/note="match to EST BG204913 (NID:g13726600)"
repeat_region 13041. .13097
/rpl_family="T-rich"
repeat_region 13075. .13199
/rpl_family="L1"
repeat_region 13200. .14202
/rpl_family="ERVK"
repeat_region 14203. .14548
/rpl_family="L1"
misc_feature 14830. .15092
/note="match to EST AA005182 (NID:g1448797) zh93h11.r1"
repeat_region 15053. .15274
/rpl_family="L2"
misc_feature 15413. .15837
/note="match to EST A1249019 (NID:g3844416) qh80h02.x1"
repeat_region 16526. .16565
/rpl_family="AT-rich"
repeat_region 16932. .17268
/rpl_family="MER1_type"
misc_feature 18420. .19031
/note="match to EST A1064746 (NID:g6359018)"
repeat_region 18715. .18876
/rpl_family="MIR"
repeat_region 18934. .18979
/rpl_family="MIR"
repeat_region 18980. .19280
/rpl_family="Alu"
repeat_region 19281. .19381
/rpl_family="MIR"
misc_feature 20122. .20258
/note="match to EST AW467230 (NID:g7037336) he07g06.x1"
misc_feature 20202. .20258
/note="match to EST AA425210 (NID:g2106002) zw47e06.r1"
repeat_region 20412. .20761
/rpl_family="MALR"
misc_feature 21074. .21452
/note="match to EST AA860276 (NID:g2954271) aj58e05.s1"
misc_feature 21234. .21392
/note="match to EST AA425210 (NID:g2106002) zw47e06.r1"
misc_feature 21234. .21392
/note="match to EST AW467230 (NID:g7037336) he07g06.x1"
repeat_region 21394. .21436
/rpl_family="AT-rich"
repeat_region 21474. .21511
/rpl_family="AT-rich"
repeat_region 21496. .21681
/rpl_family="Alu"
repeat_region 21862. .21968
/rpl_family="Alu"
repeat_region 21969. .22101
/rpl_family="Alu"
repeat_region 24521. .24613
/rpl_family="T-rich"
repeat_region 25161. .25443
/rpl_family="Alu"
repeat_region 26345. .26463
/rpl_family="L1"

Query Match 30.6%: Score 2011; DB 9; Length 160185;
Best Local Similarity 99.9%: Pred. No. 0;
Matches 2033; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
4536 GAAGATGAACAAGTTTCTGCTTTGAGGCACAGACTTTTCTGGAAGACGAGCCAC 4595

|||||
Db 55888 GAAGATGAACAAGTTTCTGCTTTGAGGCACAGACTTTTCTGGAAGACGAGCCAC 55947
Qy 4536 CTGAAGGAGAGCAGACAGAGAGCGTCGTGAGCATTTGGAGCCTTGGAACTCACATTTCTGAG 4655
|||||
Db 55948 CTGAAGGAGAGCAGACAGAGAGCGTCGTGAGCATTTGGAGCCTTGGAACTCACATTTCTGAG 56007
Qy 4656 GACGCTGGACCAAGTTTGGCTCTTCCCTGCTTTAAAGAGCATGGGGCTTTCTCCCC 4715
|||||
Db 56008 GACGGTGGACCAAGTTTGGCTCTTCCCTGCTTTAAAGAGCATGGGGCTTTCTCCCC 56067
Qy 4716 TTCTTCCTTTCCCTTTTGGATGTGAATCTGTGAAGAAATGCCCTGGCAGCTTTTCA 4775
Db 56068 TTCTTCCTTTCCCTTTTGGATGTGAATCTGTGAAGAAATGCCCTGGCAGCTTTTCA 56127
Qy 4776 CTTTGTGCTTGAATGACAGCAGTGCAGCAATCTGAGCTCCCACTGTTCCTGCCA 4835
|||||
Db 56128 CTTTGTGCTTGAATGACAGCAGTGCAGCAATCTGAGCTCCCACTGTTCCTGCCA 56187
Qy 4836 CATCACACAGTATCATTTCCAAATTCAGATCATGCACAAAGATGATTCACCTGCTG 4895
|||||
Db 56188 CATCACACAGTATCATTTCCAAATTCAGATCATGCACAAAGATGATTCACCTGCTG 56247
Qy 4896 CACTTCTCAATGCCCTGGAGAGATTTTAAATCTTCTTAAATTTAGATTTCAATCCAGTCT 4955
|||||
Db 56248 CACTTCTCAATGCCCTGGAGAGATTTTAAATCTTCTTAAATTTAGATTTCAATCCAGTCT 56307
Qy 4956 AGCAGCTGATCTCATTTGGGATATGAGAAAGCATGCCATGCACTATCTTGGGGCCTTTA 5015
Db 56308 AGCAGCTGATCTCATTTGGGATATGAGAAAGCATGCCATGCACTATCTTGGGGCCTTTA 56367
Qy 5016 ACCCACCAAGAGAGCAAGAAACCAATGAATCCTTTGATGACGTCGTCACCT 5075
Db 56368 ACCCACCAAGAGAGCAAGAAACCAATGAATCCTTTGATGACGTCGTCACCT 56427
Qy 5076 GTTTACAATGCTCCTCTTTT-AAAAAATAATGAGATTAAAGATTCTTCAGAGATTA 5134
Db 56428 GTTTACAATGCTCCTCTTTTAAAAAATAATGAGATTAAAGATTCTTCAGAGATTA 56487
Qy 5135 AATATATATCATTTTAATGATTAATGATTAATTTAAACCTTAAGTAGGGTCCAGCCT 5194
Db 56488 AATATATATCATTTTAATGATTAATGATTAATTTAAACCTTAAGTAGGGTCCAGCCT 56547
Qy 5195 GCTTCTGAAAAACCAATATGCCGACGGGTGTGCCACACCAAGAGAGGGAGAGC 5254
Db 56548 GCTTCTGAAAAACCAATATGCCGACGGGTGTGCCACACCAAGAGAGGGAGAGC 56607
Qy 5255 CTGGCTGTGACCCGTGGCTTCCCATGCTCTTGGCTCACCGGCAAGTCCCTATCT 5314
Db 56608 CTGGCTGTGACCCGTGGCTTCCCATGCTCTTGGCTCACCGGCAAGTCCCTATCT 56667
Qy 5315 GGAAGTATGAAAAATGTTAGCAATTAATPACCAAGACACCTCATCTGCTCCGCCAGTG 5374
Db 56668 GGAAGTATGAAAAATGTTAGCAATTAATPACCAAGACACCTCATCTGCTCCGCCAGTG 56727
Qy 5375 ATGGGTTTCTGTGTAACCTGTTGACATGCGCAGGGGAGGAGAACTGAGACCTTGG 5434
Db 56728 ATGGGTTTCTGTGTAACCTGTTGACATGCGCAGGGGAGGAGAACTGAGACCTTGG 56787
Qy 5435 TCCTGTCTGAGCCTTATGAGGAGAGAGCGTGCATTTGCGGATGTGCTGCTCCATTTG 5494
Db 56788 TCCTGTCTGAGCCTTATGAGGAGAGAGCGTGCATTTGCGGATGTGCTGCTCCATTTG 56847
Qy 5495 AGATGATGGAACCCCATTTTAAATTAATATTTCTTGAATTTTGTAAATTTAGAGGT 5554
Db 56848 AGATGATGGAACCCCATTTTAAATTAATATTTCTTGAATTTTGTAAATTTAGAGGT 56907
Qy 5555 GTAGGTTTGTGTTTG-TTTTTGTTTTTTTAAAGAAACATTATACTGATAC 5613
Db 56908 GTAGGTTTGTGTTTGTGTTTTTGTGTTTTTAAAGAAACATTATACTGATAC 56967
Qy 5614 ATTGAGTGAAGAGCTTGGGATGTGAGCTAAATGCCAGGTGTTATACGCTCTTTC 5673
|||||

Db 56968 ATTGCGATGAAGACACGCTTGGGATGTTGGAGCTAATGCCAGCTCTTTATAGTCTCTTTC 57027
Qy 5674 AAGACAGCCTCCCTTTATTTGATTTGGCATTTAGGGATTAACAAGCCTTTAAACGTGATAA 5733
Db 57028 AAGACAGCCTCCCTTTATTTGATTTGGCATTTAGGGATTAACAAGCCTTTAAACGTGATAA 57087
Qy 5734 AAGATCAAAAACCTGGTTAGACATGGCAGCCTTTGGCAAGCGGTAGTACCAAGACAT 5793
Db 57088 AAGATCAAAAACCTGGTTAGACATGGCAGCCTTTGGCAAGCGGTAGTACCAAGACAT 57147
Qy 5794 AACCTCAAGTGGCTTTATGAGCGCTGCATATAGAGAGCGCTTAAGTGAAGCAACATCT 5853
Db 57148 AACCTCAAGTGGCTTTATGAGCGCTGCATATAGAGAGCGCTTAAGTGAAGCAACATCT 57207
Qy 5854 GCTCAGACGTGCTATTATACCTATATGACTGAAATGAGCCCCCTCCTATTTTGTGT 5913
Db 57208 GCTCAGACGTGCTATTATACCTATATGACTGAAATGAGCCCCCTCCTATTTTGTGT 57267
Qy 5914 TGTTTTGCACAGACTCCGGAAGAGTGAAGGCTGCCAATCTGAGTACTCAAAATGTGAG 5973
Db 57268 TGTTTTGCACAGACTCCGGAAGAGTGAAGGCTGCCAATCTGAGTACTCAAAATGTGAG 57327
Qy 5974 GAACCTGCTGCTTGGATTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6033
Db 57328 GAACCTGCTGCTTGGATTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 57387
Qy 6034 AAACGTAATAGCTTCAAAATTTTAAAGTGAAGTTCGACGCTTTTTCATCTGATATCAAC 6093
Db 57388 AAACGTAATAGCTTCAAAATTTTAAAGTGAAGTTCGACGCTTTTTCATCTGATATCAAC 57447
Qy 6094 AATGTAGTCTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6153
Db 57448 AATGTAGTCTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 57507
Qy 6154 TACATTTGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6213
Db 57508 TACATTTGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 57567
Qy 6214 TATATAGCTATGAGCCTCATGTCATATGAAAGACAGAAATCTAGCTTACCACAGT 6273
Db 57568 TATATAGCTATGAGCCTCATGTCATATGAAAGACAGAAATCTAGCTTACCACAGT 57627
Qy 6274 TGCACAATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6333
Db 57628 TGCACAATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 57687
Qy 6334 GCTCTGTGATGTCAGTGCAGATGTAACAATTAACTGCTATTTCTCATACTTTTGATA 6393
Db 57688 GCTCTGTGATGTCAGTGCAGATGTAACAATTAACTGCTATTTCTCATACTTTTGATA 57747
Qy 6394 CTACTTGTACTGATGTCATTTTGAAGAGACATTTGGTGAAGTCTGATCCCTTTTGATTT 6453
Db 57748 CTACTTGTACTGATGTCATTTTGAAGAGACATTTGGTGAAGTCTGATCCCTTTTGATTT 57807
Qy 6454 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6513
Db 57808 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 57867
Qy 6514 TGTATTAGCTTTTCAATCAAGTTTGTACAAGCTGGAGAAATTAATTAATTAATTAATTAAT 6568
Db 57868 TGTATTAGCTTTTCAATCAAGTTTGTACAAGCTGGAGAAATTAATTAATTAATTAATTAAT 57922

RESULT 5
AF394782 5742 bp mRNA linear PRI 05-NOV-2001
LOCUS AF394782
DEFINITION Homo sapiens rap guanine nucleotide exchange factor mRNA, complete cds.
ACCESSION AF394782
VERSION AF394782.1 GI:15077825
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Gao, X., Satoh, T., Liao, Y., Song, C., Hu, C.D., Kariya, K. and Kataoka, T.
TITLE
Identification and characterization of RA-GEF-2, a Rap guanine nucleotide exchange factor that serves as a downstream target of M-Ras
JOURNAL
J. Biol. Chem. 276 (45), 42219-42225 (2001)
MEDLINE
21551169
PUBMED
11524421
REFERENCE
2 (bases 1 to 5742)
AUTHORS
Gao, X., Satoh, T., Liao, Y., Song, C., Hu, C., Kariya, K. and Kataoka, T.
TITLE
Direct Submission
JOURNAL
Submitted (24-JUN-2001) The Department of Molecular and Cellular Biology, Kobe University Graduate School of Medicine, 7-5-1 Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan
LOCATION/Qualifiers
1..5742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/note="RA-GEF-2"
/codon_start=1
/product="rap guanine nucleotide exchange factor"
/protein_id="AAK83368.1"
/db_xref="GI:15077825"

CDS

183..4712
/note="RA-GEF-2"

BASE COUNT 1757 a 1091 c 1323 g 1571 t
ORIGIN
Query Match 24.5%; Score 1612; DB 9; Length 5742;
Best Local Similarity 66.2%; Pred. No. 0;
Matches 2548; Conservative 0; Mismatches 1210; Indels 90; Gaps 12;
Qy 45 ACCAATGTTTCTTCATTATGAACCAACCAATCCAGCTTAACCATGAGATTATGGGC 104
Db 600 AGAAGATTTTCGAAATTAATTAAGAGGAGCGCAAAACCTTAATGATGATGGAG 659
Qy 105 CACGAGAGAAACACTTCTCTGCGAGATTTCAAAAACCTGATCTTACGACACTTC 164
Db 660 GTTAACGATATCTTCTCTCCAGCTGATCTTACCAAGATGCATCTCACAGAAACCTT 719
Qy 165 CACCCACAGGTAGACCAAGCTTTCTTACGACATTCAGAGTGAATGATCACTAGTATTC 224
Db 720 CATCCACAGGTAGCTAGTCTTCTTACGATCTGCTTGTAGACATTTGCCAGTGAATCT 779

225 GGAGACAGCAGTCTTTGATATCTACAGGCCAGAAAGCGCTGTGATATGAC 284
11 11111111 1111111111 1111 11 1111 11 1111 11
Db 780 GGAGACAGCAGTCTTATCTATATCTATCTACGGCTACGGAGAGTAGAGATAGAT 839
285 CTGAGTGGTGTCCAGAAAACAGCAGTGGATTCCGAAGACAGCAGATGAAGACATT 344
111 11 111111 11 111111 11 111111 11 1111 11
Db 840 TTGACACGCTTCCGAAGAGCCTGTGATTTCTGAGGATGACGAAGAGATGAAG 899
345 GAGAGAGCATCGATTCCTTGATGACGAGGACATTGTAGAGACTGCTTAAGAAAGAC 404
11 111111 11 111111 11 1111 1111 11 11 11 11 11
Db 900 ATTGATCGAAGATTCATTGAGGGCGAGATCTTGTTCGAAATGTCTTAAAAAGAA 959
405 CCAATTCACCGGACAGATGACATTGAAACAACCTCTTGAAATTTATGCACAGTGGCT 464
11 111 111 1111111111111111 1111 111111111111 1111
Db 960 CCTGCAGACAAAACATGATGACATTGACATTGACAAATGCTGTGATATGACACAGCTCCT 1019
465 GCTTTTGGCCAAATGACAAATGTCAAGAGGAGAACTGTGCTGTGATGGTGTTCGCA 524
11 1111 1111 1111 1111 1111 11111111 111111 11 11
Db 1020 GCATTTGCAACATGACCATGTCTGTGAAGAGAACTCTGCTCAGTGAATTTTGA 1079
525 GTGGTGGAAAGACGAGGACCATAGTGAATGATGTGAAGAGCTGACCTGTGCTA 584
11111 11 11 11 11 11 11 11111 111111 11111 1111
Db 1080 GTGGTAGACAGCGCTGAGCTATTATCTTGAAGATGGCAAGAGCTTACATCATGTGAT 1139
585 GTGATTCCTCAATGATCTGTGGAAGTGAATTCAGATGAAAAGCAGAAATTAAGTCTGC 644
11 111 11 11111111 11111111 111111111111 1111 111
Db 1140 GTTATTTTAAAGCGCAGCTGTGAAATCACTCATCCAGATGAAAAGTTTAAATTTGTTT 1199
645 ATGGGAATATGTTTGTGTCTCTCTACCATGAGCAAAAGAAATACATGAAGAGTGTATG 704
111111111111 1111 1111 1111 11111 1111 111 111
Db 1200 ATGGGAATATGTTTGTGATTAATCTCCACTCTGATGAACAGTACATGATGGAATTTGTC 1259
705 AGACAAAGTGTGATGACGCCAGTTGTCTGATAGCCAGCAGATTAAGTCCGTAT 764
11 11 11 1111 11 11111111111111111111 11 111
Db 1260 AGACATTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1319
765 CTCAATCAATGAAAGAAATGCAAAAGTTGAAGAGGAGAGATTTGTTATGTTG 824
11 11 11 1111 11 11111111 11 1111 11111 11111 1111
Db 1320 TTAACCATGTGAAAAAATACCCATTAAGTTGAGAGAGAGGAAATTTGTTATGTA 1379
825 AAGAACACCGAAGCTGATGCACTGGAACAGAAAGGACACATTTGATCAAGAGGT 884
11 11 11 1111 11 11111111 111111111111 111111
Db 1380 CATGACATCGGAGACTAGACCGAGTGAACAGAAAGACACATTTGATGATGACGA 1439
885 ACCTCAGAAAGTTAATCATGATTTGGTGAAGAGCATTCAGTAAATCAACATTC 944
11 11 11 111111 11111111 1111 11111 11 1111111 11
Db 1440 ACACCTGACGCTCATTAATGATTAATGAAGAACTTCATGATCCAACTTAT 1499
945 ATGAAGACTTCTGTGACCTATAGACTTTTCTTAGCCCAATGGAAGTGGCACA 1004
111111 1111 11 1111111111 1111 11 1111 11 11 1111 11 11
Db 1500 ATGAAGATTTTCTAATCACTTACAGACATTTCTTGAAGTCTTGTGATGTGGGATC 1559
1005 AAGTATTTGAGTGTTAATGACCCGAGCTCAGGATTAAGTTTACAGGGTAGTATTA 1064
11 111111 111111 1111 1111 11111111 11111 111111
Db 1560 AAACCTATTTGAGTGTTAATGATGATTAAGCTTAAGGTGACAGCGATTTATTA 1619
1065 TTGTGGTAAATATCACTTGAATGATGATGATGATGATGATGATGATGATGATGAT 1124
11 111111111111 1111 11111111 1111 1111 11111111 11
Db 1620 TTATGGTAAATATCACTTGAATGATGATGATGATGATGATGATGATGATGATGAT 1679
1125 GAAGAATTTGAACAATCTGGAAGAAGAAATGGTGGACACTTAAGGCTGTGAAT 1184
11 11111111 11111111 11 111 111 111 11111111
Db 1680 GAGGAATTTGAAAAAATCTGGAAGATACAAAGATTAATGCTATTCGCGTTATTAAT 1739
1185 ATGCGCTGTGCTTAAGCAAAAAGAGATGATGATGATGATGATGATGATGATGATGAT 1244
11 11111111 11 1111 11 11 11 1111 111111 11
Db 1740 ATTGCTGTGTCGAAGGCTAAGTGAGACAGGTTGTGTGCAAAAAGGCTCCCGAG 1799
1245 GCTCCTTTGCTTTATCTTACTTGAGGCTCTGAGAAAGGATTTGGAATCTTTGTTGAC 1304
11 11 11 11 11 11111 111111111111 11 11111111
Db 1800 TCCCTCTCAATTAAGCTTAATGAGGAGTGAAGAGGATTTGTTATTTTGTGAA 1859

1305 AGCTAGATTTAGTAGACAAACAACTGAAAGCAGGCTGAAAGCGGGGATCAGATATTA 1364
11111 1111111111 1111 1111 11111111 11111111 11111111
Db 1860 GAGATGAACCTGTGAGCAAAAGCTGCTGATTCAGAGCTAAACGTGTGTATGATATG 1919
1365 GAATTAATGCCCCAATCTTGAACATTCAGCTGTCAAAAGCTATGGAATTTCTTGA 1424
11111111 11111111 1111 1111 11111 1111 11111 111111
Db 1920 GAATTAATGACAAACCTTGAATATTAATTAATGAAGCCGTGAATTTTGAAG 1979
1425 AATAACACATTTATCTATCACTGTAAGAACCAATTTATTTGATTAAGACCTTCA 1484
11111 1111 11 11111111 1111 1111 1111 11111 111111
Db 1980 AATAATATCTATCTGCACTTACTGTGAAGACCAATTTTGTGTGAAGGTTACTT 2039
1485 ACAAGATTCAGAAAGAGAAAAAGAAATGCTGCCGCCACCTTCCCTAAATTTGTCATTT 1544
11 111 111 111 1111 1111 1111 1111 1111 1111 1111
Db 2040 TTTAG---GACGTAAACAAAGAAATCTGGTGTTCCTCATATTTCCAAATTTGTGTA---A 2093
1545 AAAAAGCCAGTGTCTTCCATTCAGATCTTGTGTGATGTAGAACAGGTGATAGGA 1604
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Db 2094 AAAAAGATTAATCGCCATCTATCCAGCATGTGCGAGAGATTTGACAA---GACATCA 2150
1605 CTTGAAAAGTGAACAAAAAGTAAAGCCACACTGTG---GAGAGAGAAACAACTG 1661
11 11111 11 111111 11111 11111 11111 11111 11111
Db 2151 CAGGAGAAAAGAAATGAAGAAATTAAGCAAAATCTGTTTCAAGTGAAGAAACAAATC 2210
1662 AAAAAGATCTGACAAAGCTCGATCAGTATCTTCCACAGAAACCATTAATGATATT 1721
2211 AGGAAGATTTTGTATTAACAGATTTAGTATCTTCCCTCCAAAGCTATTTAGTATGA 2270
1111 1111 11 1111 1111 1111 1111 1111 1111 1111
Db 1722 GGGATTTGTGAGTCTCAAGATGACAGCATAGTATTAAGGACGAAACATATCCA 1781
11 11 11 11111111 1111 1111 1111 1111 1111 1111
Db 2271 GGGCTAAGCAATACAAAGATGACAGCATTTGTGGGAAAGGACATGTAGCATGTG 2330
1782 ACTGCAATGCGCTGAGTGAACCTATATCAAGATTAATCTGATTTATTCAGATACAT 1841
11 1111 11 1111 111111 111111 11111 1111 1111
Db 2331 GCTATATGCCCCATCCCTGTGAACCACTCATCCAGAGGCTGTATCTCGACCTACC 2390
1842 CATGCAATTTAGACTT-----CAGTGTACTCTGTGACTTGGCAGATCAA 1886
11 11 11 11 11 1111 11 11 111111 11 11 111111
Db 2391 ACCAGATATGTTGATTTTCCAAATCCTTCAGCTGTGTTTATATATATCCTGTACAA 2450
1887 GTGCTAAGGTTTTTAAGCTGTACAGCAAAAGCGCTCATCATCATGATGAAGACAT 1946
11 111 1111 11 11111111 111111 111111 11111 1111
Db 2451 GTTATTAAGATTTTCAAAATGATGACGAAAGTGTCTACATTAATCATAGTAAGACACC 2510
1947 ACAGCAAAAGAGTGTCTATTCAGCTATCAGGAGATTTGCTGTACTGCAACCCGGAT 2006
1111 1111 1111 1111 1111 11 1111 11 11 1111 11 11
Db 2511 ACAGCTAAAGAGTATTTTTCATGCTGTTCATGAATTTGTTGACCGGTGATCCGAG 2570
2007 CAATATCAATATGTAGGCTCTGTACACACTGAGGAGTAAATCAACAAAGAACTT 2066
11111 1111 11 1111 11111 111111 11 1111 11111111
Db 2571 ACATATTTCTCTGTGAAGTTCTGTACTCTCGAGGAGTGTCAATAACAGAAAGACTT 2630
2067 CCAGATCAGCTTCCAACTTGAGACAGAAATACAACTGATGGAAGGATTAATTCGAA 2126
11111111 11111 11 1111 1111 1111 1111 111111 1111
Db 2631 CCAGATCAGCTTCCAAATTAAGCTGATGAATTAACACTCAATGAAGATTAATCTTAA 2690
2127 AACAAATGAAACAAAACTCTTTGTCAGATGAAGATGCTAAGAGTTGTGAAGAG 2186
11 111111111111 11111111 111111 1111 11 11
Db 2691 AATAACATGAAACAAAGCTTTATGATGAAGATGCTAAAGAACTAGTTAAGGAA 2750
2187 ACTCAATTTCCCTCTTCACTGACGACAGCTGTGAAGTTGCAACAGAGCTCTATGCGA 2246
11 11 11 111 111111 11 11 11 1111 11 11 1111 1111
Db 2751 AGCAGACTATCCATGCTGACGCTACATTAATGAGAGTGGCCACCAAGCTGTAAAGAG 2810
2247 AATTTGAACCTTTTGCACATTTGAACCTAGTAAATATATAGATTAATTTAAACT 2306
11111 1111 11111 111111 1111 11 1111 1111 11111 11111
Db 2811 GACTTATGATTTGTTGCTATATTTGAACGAGTGAAGATCATGATGACCTTTTAAGTTA 2870
2307 AGATCAAAAACAGCTGTGCAACCTGAAGAGATTTGAAGAGATCAATTAACGAGAAACA 2366
11 1111 11 1111 11 11111 1111 11 1111 11 1111 11 1111
Db 2871 AATTCAAAACAGGAATATCTCAATTTGAAGAGTTGTGAAGACATTTTAAACAAAGAGACA 2930
2367 TTTTGGTAGCATCTGAAATTTTCAGAGAAACAAACAGCTGAAGAGATGAAGATCAT 2426

Db 2931 TTCTGGGTGCTCCAGAAATTTTAAGTGAAGCAATCAGCTCAAGCAATGAAGATATT 2990
QY 2427 AAGCATTTTCAGAAATAGCACTGCACTAGAGGAATGCAAGATTTTAACTCAATGTTT 2486
Db 2991 AAGCATTTTAAATTTGCACTTCAATGTGCGAAGATGTAAGAACTTCAATTCATGTTT 3050
QY 2487 GCATATCATGTCGCTTAAACCTGGCAGCAGTGGCAAGACTGGCAAGCAGCTGGAGAAA 2546
Db 3051 GCATATTAATGTCGCTTGAACCTGGCATCTGTAGCAAGACTGAGAGAACTGGAGAAA 3110
QY 2547 CTTCCTCAATTAATAGCAAAACATATTTCAAGATCTCCAGACCTGTTGATCTTCACA 2606
Db 3111 TTACCAAGCAAAATACAGAAACATCTTCAAGATCTACAAAGCACTTTTGTATCATAGA 3170
QY 2607 AACATGGCAAAATATGTAATGTTTCATAGTCAAAATCTACAACTCCCATATCCCT 2666
Db 3171 AACATGGCAAAATATGTAATGTTTGTAGTCAAAAGTATGAGCAATTCATATTC 3230
QY 2667 CTATTCCTCAATTTCAAAAGATCTCACCCTTCCTCAGCAAGAAATGATCAAAAGTA 2726
Db 3231 CTCTTCCTGTTGTCAAGAAAGATGACATTTCTACATGAAGAAATGATCCAAAGTA 3290
QY 2727 GAGGCGTGTGCTCAATTTGAGAGCTAAGGATGATTGCAAAAGAAATTCCTCAGTTGGC 2786
Db 3291 GATGCTTTTAACTTAAGCTTTGAGAGTTAAGATGATTTCCAAAGAAATCCGCCAAGTTGT 3350
QY 2787 CGAATGGCTTCACTGACATGACCTGCCCTCATGTTTCAGAGCTGGAGAAAGAAATG 2846
Db 3351 CGAATGCTTCTCTAAGACGACCCAGCTATGATTTTGACAGAGAAAGAGTGG 3410
QY 2847 CGAGTTTGGGGTCTCTCACCAGGGTACTACAAATGCAACAGTGAATGTTGCTCAG 2906
Db 3411 CGAGTCTGGGTCTCAGTCAAGAGCAAAATTCAAACATGCTGAGTGT--TCAG 3467
QY 2907 ACAGTGTGATTAATAAAGCGGATACGCTAGTTCCTTTCTCAATGCCAAAGAACTTTAT 2966
Db 3468 GAGAGTCTCACAAAAGGAGGAGCGGCGAGCTCTGCTTAATGCAAGAAAGCTATAT 3527
QY 2967 GAAGATGCCAAATGCTCGAAGAGTGAAGCAAGTACCTTTCCATTTTGAAGTGAATG 3026
Db 3528 GAGATGCCCAAAATGCGCAAGAGGAGTGAAGCAAGTATCTTCCAGTCTCGATGTAAGACA 3587
QY 3027 GAGAGAGAGTCTTCAGACATATCTCTGAGTGAAGGCAAGCAACATGTCCT 3086
Db 3588 GATGAGGAGAACTTCGATGATGTCACTTACATGGAGGCTGCTATGTTGATCTTACC 3647
QY 3087 AAGATCTGTGACAAAGAGCTGTCA--ATCCGAGACTCTCTCCAGTACCTCAAG 3143
Db 3648 AAGAAATTAAGTGAAGAAAGATCAGCAAAATCTGAATGTCTCCAGTGCCTATGAG 3707
QY 3144 GCAGGCTCAACAGAAAGCTCACTGCTGCGACAGGCCAGCAGCAGCAGCAGCAGCA 3203
Db 3708 TCACCTGGCCCAACAACTAAAGCCCACTTCAATCAACCC----- 3746
QY 3204 CATTAATCAACAGGAGTACAGTTCAGGCGGTGCTCCCTTATTCCTTCAGGAAGAAA 3263
Db 3747 CACGAGATTAAGCCAGGTGCTTCAAGTGCAGCTGTATTTTGAAGCCCATCAGSAGAAAG 3806
QY 3264 GTGCGCTTAAGATCTCCACCTTTTGGCATAAATCTCCACAAGCTTTTAAAAAAATT 3323
Db 3807 GAGCAAAACAAAGACCTGCACTGAATACAAAGTTTACCTCAGAAAGTTTGAAGAA-- 3864
QY 3324 CTTTCTTGTCTGAAGAGAGTGTGGAAGCTACAAAGAAAGGCTGAAGATACATA 3383
Db 3865 -----CTGAAGAAATTA-----GTGGTAAAGAGCTATACAGAAACACATATT 3905
QY 3384 TCMAATGATCTTGGCAGCTTCTTCCTCCCTACTCTCCACAGAGATTCTCCAAAGAAA 3443
Db 3906 TCTGTGGCGTCAATTTTACATTTCTAGTCTCTGCAATCTCTCAAGGCTCCCTCACAATA 3965
QY 3444 GGCTATATCTTGGCTCCAGTGTACTGTGATTAATTTTTCAGATCTGTGCTCACACTGAA 3503

Db 3966 GGTACACACTTATTCATCAGCTAAATCTGACAACCTTGCTGACTCCAGCATAGTAG 4025
QY 3504 ATTTCTTCACGATTCAGATTTTGTAGCAATTCGCTTTTGTAGTGCAGTCTCACTG 3563
Db 4026 ATTTCTTCACGATTCAGATTCGATTTGTGTGTGATCTCATCTGCTCACTCTA 4085
QY 3564 CAGATGAGAGGCGCAGAGGCAATTCCTGACAGTCTGGAAGCAAAACCTTAGGATGGC 3623
Db 4086 CAGATGAACGGTGTCTCTCAGGCCCCCTGGCAGTCCGATTCACATGCGGCAATTGGA 4145
QY 3624 AGATGAGAGGCGGAGCATGATTAACCTGATCAGTATAGCTTGGGCTTATGCACCA 3683
Db 4146 AAGACAGACGACCTTCAGAGATAGAGATCATATGTAACATAGCCCTGGGTGACACTC 4205
QY 3684 ATGTCCGAGGCGGAGCTTATATGTTACAGCTACAGTATTTCTTCCACAGCAGCAG 3743
Db 4206 TTGAAGCCATCTCTAATCAAGTGTGATGCTCTCA-----TCGTCGTGAGCAATGA 4259
QY 3744 GAATTTCCAGAGATCAGGGGATCGCGCTCATTGATGCTGTGACAGTGGCCGTGGG 3803
Db 4260 GAGATTTCTCAGAGCATATCATTT-----ATGAAAGCAGCTGACAGTGTCTGGA 4310
QY 3804 AGCTGAGCTCATGCTCAGAGTGGCTCCATGATATATACAGACATCCAGCAGAGA 3863
Db 4311 AGTTGCACTTCTGTGTCAGACAGCTCCCATGACAACTTCCAAAGCTTCCAAACCAANA 4370
QY 3864 AGCTGGGA 3871
Db 4371 AGCTGGGA 4378

RESULT 6
AF478567 6390 bp mRNA linear PRI 02-MAY-2002
LOCUS AF478567
DEFINITION Homo sapiens PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2 mRNA, complete cds.
ACCESSION AF478567
VERSION AF478567.1 GI:20386205
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 6390)
AUTHORS Kozlov,G., Banville,D., Gehring,K. and Ekkel,I.
TITLE Mammalia: Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL A PDZ domain containing guanine factor (GEF) interacts with the second PDZ domain of human PRLP
TITLE Unpublished
REFERENCE 2 (bases 1 to 6390)
AUTHORS Kozlov,G., Banville,D., Gehring,K. and Ekkel,I.
TITLE Solution structure of the PDZ2 domain from cytosolic human phosphatase hPRLP complexed with a peptide reveals contribution of the beta2-beta3 loop to PDZ domain-ligand interactions
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 6390)
AUTHORS Banville,D.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2002) Biotechnology Research Institute, 6100 Royalmount, Montreal, Quebec H4P 2R2, Canada
FEATURES
source
1..6390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31.1"
207..5012
/codon_start=1
/product="PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2"
/protein_id="AA021637.1"
/db_xref="GI:20386206"
/translation="MNSPVDPGARQALRKPPERPEDLNTLYSLHGMETLSNLRH QRLMSRARIERYSGNOVLFCESETIARCMYLLSGSYLVKGSWVLPDPSFGKGGK

RGCDLVLEPSEMIVENAKNDNEDSILOREIPAROSRRFRKINXGERQITDDEV
NSYLSIPADLTFRMLTENDPHQVTVHSSOGCSIASDSGSSLSIDIQATSEVGDV
DUTRLPEPVDSEDEDEEDIEDRTDHPLOGDIVRECELEKREKPADDDIDBOLLEFMH
OLPAPANTMSVRELCSVMIFEVENQAGAIILBEGOLDSMYVILNKTVEISHPDK
VENIFPMSFEITITPLDKOYMHGIVRTKVDQCFVCIQOQDRTWRLNTEVENHTHVEE
EGEIVMHEHEHRELBDSGRKHIVIKATPERILMLIEHSIVDPTIEEDLILYRTF
LESPLDVGIKLLEWFKIDSLDKVTRIVLLVWNNHFNEDPANTRLLEFEKLEED
TKMNGHLKLNLINACAKAKMROVYLOKRSFNSFANGSEKGFJFVGEVGEKSD
AADGLKRGDOMEYNGONFENITFMKVEILRNHTALVTKTNLIPFKELIPEFTEO
EKSGYPIHPKIAEKSNHSTONHPGDIUEQISOEKSQKVVANITYSGRRNKTRLDK
TRFSLPPLKFLSDGLOSODSVIGTHCHRSUAIMPIGTLSSSSDPLQPTTSM
DFSNSDIPDOVIRVFKVDOQSCYIISKQDTAKVYVHAHEFGLTADSLYSLCEV
SYTEGVIKORRLPDQFSKLADRIOLNRYEILKNMEETLSCDEBDQAEKYSOLEM
LOLSTIEVATOLSMDFPLFNIEPTEYIDDLFNKSNSTGLIKREFDIQNOETFW
ASELITEANOLKRMKLTIKFKIALHCECEKNFMFXIISGLINLASARILNPEKEL
PSKTEKHLQDIDITDPSRNMKTRNIISSOSMDPILPLPVAKDMTFLHEGNDK
VDGLVNEFLKMLISKEIROVAVMTSANDPAMPEMORLSOGSTSNMLDVQGAHAK
RARRSLNANKLIEDAOMARKVOYLSLVEVDEBEKQOMSLQMEPAVGTLTANLS
EKRSKSSSEMPVPMRSAGQTTKAHLHQPHRSOVLQVPAVNLHPIRKKQTKDPALN
TSLPKVLTTEIEISGKHTEEDTISVASSLHSSPASPQSGPHKGYTLIPSAKSNLS
DSHSEISRSISVNCVSDMSAALODERCSOALANPESTGALEREKHEASGIDHS
OHGPMWTLKPSLKLCLAVSSSVSNEEISQEHITIEADSGSGWTSKSSSHDNFOS
LNPBKSDFLNSYRTHLDPTAEVEPDSPEYSCSKSRGCKGKSLERKSWTSS
SSLSDTYEPNGTVKRVLESTPAESSEGLPKADPDVYKTVSTESTERGLIVCYTS
PKKDRYREPPPTPGYGLISLADLKEGPHILKPPDYSVAORSKMMHNSLSRLPAP
SLSSNLVACVPSKIVTOPORHNLQFPHKLDGVDADSEADENEGVANSV"

BASE COUNT 1950 a 1242 c 1447 g 1751 t
ORIGIN

Query Match 24.1% Score 1585.4; DB 9; Length 6390;
Best Local Similarity 66.0%; Pred. No. 0;
Matches 2528; Conservative 0; Mismatches 1206; Indels 99; Gaps 12;

QY 45 ACCATATGTTCTTCTTATGAACACACTAGCAATCCAGCTRACCATGAGTTATGGGC 104
DB 624 AGAAGATTTTCGAAATTTACTATAAGAGAGCCGCAACATTAATGATGAGTGAG 683
QY 105 CAGCAGAGAAACACTCATCTCCTGCAGATTTACAAAACCTCATCTCTGACAGTCTC 164
DB 684 GTTACACGATCTCTCTCTCCAGCTGATCTTACCAAGATGATCTTCACAGAAAACCT 743
QY 165 CACCCACAGGTGACCCACGCTTTCTTCTAGCCATTCAGATGATGATCACTAGTATCT 224
DB 744 CATCCACAGGTGACATCATGCTTCTTCTAGCTGATGCTGTTGATGATGCCAGTCTCT 803
QY 225 GGGAGCAGCAGCTTCTCTATCTACAGGCCACAGAAAGCAGGCTGTGATATGAC 284
DB 804 GGAAGCAGCAGTTATCTGATATCTATCAGGCTACGAGAGTGAGGATGTAGAT 863
QY 285 CTGAGTGGTGGCAAAACAGCAGATTCGAGAGAGCAGCAGCATGAAGAAGCATT 344
DB 864 TTGACACGCTTCCAGAAAGACCTGTTGATTCGAGAGATGAGAAAGAGAGAG 923
QY 345 GAGAGAGCATCAGATCTCTGATGAGCAGGACATTCGTGAGAGACTGCTTAGAGAGAC 404
DB 924 ATTGATCGAAGATCATTTGAGGGGCGAGATCTTTGAGAGAATGCTTGAAAAAGAA 983
QY 405 CCAATTTGACCGACAGATGATGACATTGAACAATCTTGGAATTTATGACACAGTTGCT 464
DB 984 CCTGACGAGAAAATGATGATGACATTGGAACAATTCGTGAGTTATGACACAGCTCCT 1043
QY 465 GCTTTTGCAATATGACATGAGTGAAGGAGGAGACTCTGAGCTGTGAGTGGTTCGA 524
DB 1044 GCATTTGCAACATGACATGCTGTGAAGAGAGAACTGCTCACTGATGATGATTTGAA 1103
QY 525 GTGTGGAGAAAGACAGGACCATAGTGAATGATGATGATGATGATGATGATGATGAT 584
DB 1104 GTGTGAGAGAGGCTGAGCTATATTCTTGAAGATGGGCAAGAGCTTGACTATGATGAT 1163
QY 585 GTGATTTCTCAATGATCTGTGAAGTGAATTCAGATGAGAGAGCAGAAATACTGTGC 644
DB 1164 GTTATTTTAAAGCGCAGTGTGGAATCACTCATCCAGATGAGAAAGTGAATTTGTTT 1223

QY 645 ATGGGAATAGTTTGGTGTCTCTCTACCATGACCAAGAAATACATGAAGAGATGATG 704
DB 1224 ATGGGAATAGTTTGGATTTGATTTACTCTCCACTCTGTGATTAACAGATCACTGATGAT 1283
QY 705 AGAACCAAGGTGATGACTGTCAGATTTGTCTGCATAGCCACAGACATTAAGTCCGTAT 764
DB 1284 AGCAGTAAAGTGAATGATGATGCTGAGTTGTCTGCATAGCCACAGCAAGATTAAGGAAT 1343
QY 765 CTCATCAAGTGAAGAAAGCAATGCAAAAAAGTTGAAGAGGAAGAGATGTTGATAGTG 824
DB 1344 TTAAACCATGTGGAAGAAAAATACCCATAAGTTGAGGAAGAGGAAGAAATTTGTAAGTA 1403
QY 825 AAAGAACCCGAGAACCTTATGCAACTGACACAAAGAGGACACATTTGCATCAAGGGT 884
DB 1404 CATGACATCGGAGACATACACCGGAGTGAACACAGAAAGACACATTTGATCAAGGCA 1463
QY 885 ACCTCGAAGAGGTTAACATGCAATTTGGTGAAGAGACATTCAGTATGATCAACATTC 944
DB 1464 AACCTGAGACGCTCTCATATGCAATTTAATGAAGAAACATTCATGCTGATCCAACTTAT 1523
QY 945 ATGAGACATTTCTGTGACCTATAGACTTTTCTTTACCCCAATGCAAGTGGCCAAA 1004
DB 1524 ATGAGAGATTTTCTATTAACCTTAACAGACATTTCTGAAAAGTCTTGTGATGGGATC 1583
QY 1005 AAGTTATGAGAGGTTTATGACCCGAGCTCAGGATTAAGTTACAGCGGATGATTA 1064
DB 1584 AAACCTATTTGGAATGTTTAAAGATCGACAGCTTAAGAGATAGGTGACACGAGATTA 1643
QY 1065 TTGTGGTAAATTAATCACTTCATGACTTTGAAGAGATCTGCAATGACTGATTTTAA 1124
DB 1644 TTATGCTAAATATATCATTTTAAATGATTTGAAGGTGACCCCTGTAATGACTGATTTCA 1703
QY 1125 GAAGATTTGAAACATCTGGAAGAGAAATGGGTGAGACACTTAAGGCTTTGAAT 1184
DB 1704 GAGAGATTTGAAAAAAATCTGGAAGATPACAAAGATTAATGCTATCTCGGATTAAT 1763
QY 1185 ATCCGCTGTGCTGCTAAACCAAAAAAGAGATGATGACGTTAAACAAACATCCAGAA 1244
DB 1764 ATTGCTGTGCTGCTCAAAAGCTAAGTGAGAGACAGTTGTCTGCAAAAAGCTTCCCGAG 1823
QY 1245 GCTCTTTCCTTTTATCTTACTTGAGAGCTCTGAGAAAGGATTTGGAATCTTTGTTGAC 1304
DB 1824 TCCCTCTCAATTAATCAAGCTTAATGAGAGGAGTGAAGAGGATTTGTTGTTGAA 1883
QY 1305 AGTGTAGATTCAGATGCAAGACACTGAAGCAGCTTGAAACGGGGGATCGATATTA 1364
DB 1884 GAGTGAACCTGTGTGCAAAAGCTGCTGATTCAGACTGTAAGCTGTGATGATTAATG 1943
QY 1365 GAAGTAAATGGCCAAACATTTGAACATTCAGCTGCAAAAGCTTAAGAAATTTTGA 1424
DB 1944 GAAGTAAATGGCAAAACATTTGGAATATTTACATTTATGAAACCCGTTGAAATTTGAG 2003
QY 1425 AATTAACACATTTATCTATCACTGTGAACCAATTTATTTATTTAAAGAACTTTTA 1484
DB 2004 AATTAATCACTATCTGCACTTACTGTGAAGACCAACATTTTGTGTTCAAAAGGTTACT 2063
QY 1485 ACAAGATTTGACAGAGAAAGAAATGTTGCCCCACCTCTCTAATTTGTTGACATT 1544
DB 2064 TTATG---GACTGAACAAAGAAATCTGGTGTCTCATATTTCCCAAAATTTGCGAAAA 2120
QY 1545 AAAAAGGCGAGCTGATCCATTCAGATCTGCTGTGATGATTAACAGAGTGAATGGA 1604
DB 2121 AAAAGT---AATGCAATTTCTATCCAGATGTGCGCAGAGATATTTGAAC---GACATCA 2174
QY 1605 CTTGAAAGAGTGAACAAAAAAGTAAAGCCACACCTGTG---GGAGAGAGAAACAGCTG 1661
DB 2175 CAGGAGAAAGGAAGTGAAGAAAGTTAAAGCAATACGTTCAGTGGAGAAACAAATC 2234
QY 1662 AAAAATACCTGACAAAGACTCGGATCAGTATCTGCCACAGAAACATTAATGATTA 1721
DB 2235 AGGAAGATTTTGAATAAACACGATTTAGATCTTCCCTCCCAAGAGCTATTTAGTATGGA 2294
QY 1722 GGGATTTGGTCACTCTCAAGATGACAGCATTAATGATTAAGGACGAAAGCAATCCCA 1781

LOCUS AF117947 6401 bp mRNA linear PRI 01-JAN-2000
DEFINITION Homo sapiens PDZ domain-containing guanine nucleotide exchange factor 1 mRNA, complete cds.
ACCESSION AF117947
VERSION AF117947.1 GI:6650765
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 6401)
AUTHORS Kawasaki, H., Chen, E.J., Springett, G.M., Graybiel, A.M. and Housman, D.E.
TITLE A new family of Rap guanine nucleotide exchange factors
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 6401)
Housman, D.E.
Kawasaki, H., Chen, E.J., Springett, G.M., Graybiel, A.M. and
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1999) Center for Cancer Research/Dept. Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
FEATURES
SOURCE Location/Qualifiers
1. 6401
/organism="Homo sapiens"
/db_xref="taxon:9606"
70. 3684
/note="PDZ-GEF1"
/codon_start=1
/product="PDZ domain-containing guanine nucleotide exchange factor 1"
/protein_id="AAF22004.1"
/db_xref="GI:6650766"
/translation="MAPTLFQKFSKRTGLGAPGDARDPDCGSEMPLEDPDSQRL IYQDCERGRNVLEDSVSKRNEEDISVSDNTITSTYLGHEMTLSNLEHOLRLSAR ANTERYSQNVLFCESTLARCWYILLSGVLVKPGSGVQFGKRGKDCDLVLPSEMIYVENAANDNDSILQREIPAROSRRRKINYGKREITIDDEVNSYLPA DLTKHLNLEPHQVTHYSSQSGSISADSGSSSLSDIYQATESEVDVLTIRPEG PVSDDEDEDEIDRTDPLQGRDLVRGCEKPEKDDDDIOLLEEMHOLPAPAM TMSVRELCVWTFEVEVOAGAILLEDOQELDSVYIINGYEISHPGKVENLEPMN SFGITPTLDKQIMHIVRTKYDDCOFVCIAOODIWRILNHEKNTKKEEBSLIVH EHRELDKSGTRKGHIVIAKPERLIMHLEHSIVDPYIEDFLITRTFLESPLDVG IKLEMFKIDSLRDQVTRIVLWNNHFNDESGDPAWTRFLEEFKNLEDTYMGHLR LILNACAKAKARQVYLOKASRESPLQSLNGSKGGEIYEVEGPEKSKADSGIKR GDOIWEVGNONFENTPMKAVELLNNHNLALTVTNIFVEKELLFRTEQSGVPHI PTKAKKSNRHSIOHVPDIDIEPTSEKSKYKATVYGGRRKIKILDKTRFSILPP KLFSDGGLSQSDDSITVGRHRHSLAIMPIDGTLSSSPDLQPTSMLEDSNSDI PDQVTRVERKVDQSCYIISKDTAKKEVFAHVEGLTGADYSLCEVSVTPGVI KQRLPDPFSKLADRIQLNGRYLLKNMTEETLCSDEDAQELKESQSLMQLSTIEV ATQLSMRDPLENIEPTVEYIDDLFKLSKNGTILKFEEDIVNGETFWASIELTEA NOLKRMKLIKHFIKIALHCRCGNKNSMFAIISGILASVARIKRTWELPKSKYKHL NODLIDIPDSRMKAVYNTLSQSMOPITLPEPVYKDMFTLHSGNLSQVLYNFE KLRMISKEIRQVRYNTSNMIDPAMFROBSISQGSTNSMMLDVCGAKKRRKRSLL NAKKLIEDAQMARRVKOYLLSLDVELDEKFOUMSLQWEPAYGLTKNLSEKRSKII "

BASE COUNT 1971 a 1236 c 1435 g 1759 t
ORIGIN

Query Match 24.1%; Score 1584.4; DB 9; Length 6401;
Best Local Similarity 65.9%; Pred. No. 0;
Matches 2528; Conservative 0; Mismatches 1206; Indels 100; Gaps 12;

QY 225 GGAGCAGCAGCTCTTCTGATATCTACAGGCACGAAAGCGGCTGTGATGAC 284
11 11111111 1111111111 1111 11 1111 111111 11
DB 817 GGAAGCAGCAGTCTTATCTATATCTACAGGCTACGAGAGTGAGGTAGCATT 876
QY 285 CTGAGTGGGTGCGAAGAACAGCAGTGGATTCCGAAGACAGCAGCATAGACATT 344
111 1 111111 1 11111111 1 111111 111111 111111 11
DB 877 TTGACACGCTTCCGAAAGAGCCTGTTGATTCGAGAGATGACAGAGAAAGTAAAG 936
QY 345 GAGAGACATCAGATCTCTGATGAGCAGGACATTTGTGAGAGACTGCTAGAGAGAC 404
11111111 11111111 11 1111 1111 1111 1111 1111 11
DB 937 ATTGATCGAAGATCCATTGACGGGCGAGATCTTGTGAGAAATGCTTGAAGAA 996
QY 405 CCAATTGACGGGACAGATGATGATGATGACAAACCTGTGAATTTGACACAGTGGCT 464
11 111 11111111111111111111 1111 1111111111111111 1111
DB 997 CCTCGAGACAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1056
QY 465 GCTTTTCCCAATATGCAATATGATGATGATGATGATGATGATGATGATGATGAT 524
11 11111 111111 1111 11111111 111 11111111 111111 1111 11
DB 1057 GCATTTTGCAAACATGACATGCTGTGTAAGAGAGAACTGCTGATGATGATTTTGA 1116
QY 525 GTGGTGAAGAGCAGGACCATAGTGAATGATGATGATGATGATGATGATGATGAT 584
111111 11 11 11 11 11 1111 1111111 1111 1111
DB 1117 GTGTAGACAGCAGCTGAGCTATTATTCTTGAAGATGGGCAAGAGCTGATCATG 1176
QY 585 GTGATCTCAATGATCTGTGGAGTGAATTCATGATGATGATGATGATGATGATGAT 644
11 111 11 11 11111111 11 111111111111111111 1111 111
DB 1177 GTTATTTTAAAGCGCAGCTGTGAAATCAGTCCAGATGAGAAATTTGAAATTT 1236
QY 645 ATGGGAATAGTTTGTGTCTCTCTACCATGAGCAAGAAATATGAAAGAGTATG 704
11111111111111111111111111 1111 1111 1111 1111 1111 1111
DB 1237 ATGGGAATAGTTTGTGAAATACCTCCACCTGATGATGATGATGATGATGATGAT 1296
QY 705 AGACAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 764
11 11 11 11 11 11 11111111111111111111111111 1111 1111
DB 1297 AGGACTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1356
QY 765 CTCATCAAGTATGAAAGAAATGCAAAAGTGAAGAGAGAGAGATGATGATGAT 824
11 11 11 11 11111 11 11111111 11 111111111111111111
DB 1357 TTAACCATGTGAAAGAAATACCATTAAGTGAAGAGAGAGAGATGATGATGAT 1416
QY 825 AAGAAGCAGCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 884
11 11 11 11 1111 11 111111 11 11111111 11111111 111111
DB 1417 CATGACATCGGAACTATGACGGAATGACAGCAGGAAAGACATTTGATCAAGGCA 1476
QY 885 ACCTCAAGATTAACAATGATTTGGTGAAGAGACATTCAGTATGATCAACATTC 944
11 11 11 11 11111111 1111 1111 1111 11 111111 11
DB 1477 ACACCTGACGCTCATTAATGATTAAGAAACATTCCTGATCCAACTTAT 1536
QY 945 ATGAGACATTTCTGTGACCTTATGACCTTCTTCTAGCCCATGGAAGGCGCA 1004
11111111 1111 11111111 1111 1111 1111 1111 11 11 11
DB 1537 ATGAAAGATTTCTATTAACTTACAGGACATTTCTTGAAGTCTTTGGATGGGATC 1596
QY 1005 AAGTATGAGATGTTAATGACCCGAGCCTGAGGATTAAGTTACAGGGTATGATTA 1064
11 11111111 11111111 11 11 11 111111 1111 111111
DB 1597 AAATATGGAATGTTAATGATGATGATGATGATGATGATGATGATGATGATTA 1656
QY 1065 TTGTGGTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1124
11 11111111111111 1111111111 11111111 1111 1111111111 11
DB 1657 TTATGGGTAAATATGATGATGATGATGATGATGATGATGATGATGATGATTA 1716
QY 1125 GAGAATTTGAAAGCAATGCGAAGAGAGAAATGGGGAACACCTAAGGCTTGAAT 1184
11 11111111111111 11111111 11 1111 1111 1111 111111
DB 1717 GAGGAATTTGAAAGCAATGCGAAGATACAAAGATTAATGATGATGATGATGATTA 1776
QY 1185 ATGCGGTGCTGCTAATGCAAGAAAGAGATGATGATGATGATGATGATGATGAT 1244
11 11111111111111 11 1111 11 11 11 11 1111 111111 11
DB 1777 ATGCGGTGCTGCTAATGCAAGGATGATGATGATGATGATGATGATGATGATGAT 1836
QY 1245 GCTGCTTGTGCTTATCTTACTTGAAGCTCTGAGAAAGGATTTGGAATCTTTGTTGAC 1304
11 1111 11 11 11 111111 111111111111 11 11111111
DB 1837 TCCCTCTCAATTAAGCCTTAATGAGAGAGTGAAGAAAGGATTTGTTTGTGAA 1896

QY	1305	ACGTGATGTTACAGTACGCAACCACTGACAGACGCTTGAACGGGGGGATTCAGATATTA	1364
Db	1897	GGAGTAAACCTGGTGVGCAAAACCTGCTGATTTAGGACTGAAACCTGGTGATCGAATTATG	1956
QY	1365	GAAGTAATGGCCAAACCTTTGAAAACATTCAGCTGCTCAAAAGCTATGGAATTCCTTAGA	1424
Db	1957	GAAGTAAATGCAAAACCTTGGAAATATATACATTTATGAAAGCCGTTGAAATTTTGAGG	2016
QY	1425	AATTAACACACATTTATCTATCACTGTGAAACCAATTTATTTGATATTTAAAGAACTCTA	1484
Db	2017	AATATATCTATCTTCCACACTACTGTGAAGACCAACATTTTGTGTCAAAAGATTAATCTT	2076
QY	1485	ACAGATGTTCAGAGAGAAAGAAATGGTGGCCCCCACTCTCTAAATTTGGTGACATT	1544
Db	2077	TTTNG---GACTGAACAAGAGAAATGTGGTGTCTCTCATATTTCCCAAAATTTGCTGAAAAA	2133
QY	1545	AAAAAGCCAGTGGCTACTCCATTCAGATCTTGCTGTATGATGTGAACAGGTGATAGGA	1604
Db	2134	AAAAGT---AATGGCAATTTCTATTCAGCATGTGGCCAGAGAAATTTGAACA---GACATCA	2187
QY	1605	CTTGAAAAAGTGAACAAAAAAACTAAAGCCAACCTGTG---GGAGGAGGAAACACCTG	1661
Db	2188	CAGAGAAAGAGATGAAGAAATTAAAGCAAAATACTGTTTTCAGGTGGAGAAAGAAACAAATC	2247
QY	1662	AAAAAGATACTCGACAGAGACTCGGATCACTATCTTGCCACAGAAACCATACATGATATT	1721
Db	2248	AGGAAGATTTTGGATTAACACAGATTTAGTATCTTGCCCTCAACATCTATTAGTGATGA	2307
QY	1722	GGGATTTGTCAGTCTCAAGATGACACCATATATGATATTAAGGCAACAAACACATCCCA	1781
Db	2308	GGCTTAAGCCAAATCCAAAGATGACACATTTGGTGGAAACAGACACTGTAGGCATATGCTG	2367
QY	1782	ACTGCATTTGCTGTCACTGGAACCTTATATCCAGTAAATCCGATTTTATTCAGTCAAT	1841
Db	2368	GCTATTAATGCCAATCCCTGGAAACGTCATCCAGAGGCCCTGATCTCTCCAGCCTTACC	2427
QY	1842	CATGCAATTTTAGACTTCACTGCTACTCTGACCTGCCAGATCCAGTCACTGCTGAAGGTTT	1901
Db	2428	ACCAATATGTGGATTTTTCATTCCTTGAGATATCCCTGATCAAGATATTAAGATTTTC	2487
QY	1902	AAGCTGTATCGACAAACCCCTACATCATGATCATAGTAGAGACACTACAGCAAGGAAGTG	1961
Db	2488	AAAGTGATCGCAAAAGTCTCTACATTTATCATCATGATTAAGAACACCACAGCTAAAGAAATGA	2547
QY	1962	GTCATTCAGGCTATCAGGAGATTTGCTGTTACTGCCACCCGGGATCAATATTCGCTATGT	2021
Db	2548	GTTTTCATGCTGTTCATGAAATTTGGTTGGTCCGGTGATCCGCAACATATTTCTCTGT	2607
QY	2022	GAGGCTCTGTCAACACTGAGGGAGATTAATCAAAACAAGAGAACTTCCAGATCAGCTTTC	2081
Db	2608	GAAGTTTCTGTTACTCTGAGGGGTGTCATTAACAGAGAAGAACTTCAAGATCAAGTTTCC	2667
QY	2082	AAACTTGCAGACAGATTAACAATCTGAGTGAAGTATTTATCTGAACCAACATGGAACA	2141
Db	2668	AAATATAGCTGTATGAATTAACATCACTCAATGGAAGTATTAATCTTAAAAAATACATGGAACA	2727
QY	2142	GAAACTCTTTGTAGATGGAAGATGCTCAGGAATTTGTAGAGAGAGTCAATTTCCCTC	2201
Db	2728	GAAACCTTATGTGAGATGAAGATGCTCAAGAACTAGTTTAAGGAAAGCCAGCATATCCATG	2787
QY	2202	CTTCAGCTCAGACACTGTGGGAAGTTGCAACACACCTCTCATGCGAAATTTTGAACCTTT	2261
Db	2788	CTGCAAGCTCATACCATTTGAGGTGGCCACCACAGCTGTCAATGAGGGACTTTGATTTGTTT	2847
QY	2262	CGCAACATTTGAACCTACTGAATATATATGATGATTTTATTAACCTCAGATCAAAAACCGC	2321
Db	2848	CGTATATTTGAACCGACTAGATATCATGATGACCTTTTAAGTTTAATTCGCAAAAACAGA	2907
QY	2322	TGTGCCAATCGGAAGAGATTTGAAGAGTATTAACCGAAGAAACATTTGGTATACATCT	2381
Db	2908	AATACTCATTTGAAAGAGTTTGAAGACATTTGTAAACCAAGACATCTTGGGTTTGCCCA	2967
QY	2382	GAATATTCAGAGAAACAACACCACTGAAGAGATGAAGATCATTTAAGCATTTTCATCAAG	2441

Db	2568	GAATTTTAACTAGCAACCAATAGCTCAACAGCATGAGATTATTAACGATTTTATTA	3027
Qy	2442	ATGACGCTGACGTGTAGGGAATGCAAGAAATTTAACTCAATGTTTGCATCATCAGGCG	2501
Db	3028	ATTGCACTTATTTGTGAGAAATGTAAGAACTTCAATTCATGTTGGCAATATATAGTGGC	3087
Qy	2502	CTAAACCTGGCCACCATGGGCAAGACTGCGAAGGACCTGGGAGAACTTCCCAATTAATAC	2561
Db	3088	TTGAACTGGGCACTCTGAGCAAGACTCAGAGGAACTTGGGAAAGATTGCCAAGCAATATC	3147
Qy	2562	GAANAACATTTACAGATCTCCAGAACCTGTTGATCTTCCAGAACATGSCAAATAT	2621
Db	3148	GAGAAACATCTTCAAGATCTACAGACATTTTGTGATCCATCTAGAAACATGSCAAAGAT	3207
Qy	2622	CGTAATCTTCTCAATAGTCAAAATTCACAACTCCCATATCCCTATTCCTCCAGTTATC	2681
Db	3208	AGAAATTTCTTTGTGTACTCAAGATATGAGGCTCCATATTTCCACTCTCCCTGTTGTC	3267
Qy	2682	AAAAAGATCTCACTCTCTTCCAGGAAGAAATGACTCAAAAGTAGAGGCGCTGGTCAT	2741
Db	3268	AAGAAAGATATGACATTTTCTACATGGAAGAAATGACTCCAAAGTAGAGTTTATGTAAC	3327
Qy	2742	TTTGAAGAGCTAAGATGATGCAAAAGAAATTCGTCACGTTGGCGGAAATGGCTACGTG	2801
Db	3328	TTTAGAAGATTAAACAAATGATTTCCAAAGAAATCCGCAAGTTGTTGGAATGACTTGTCT	3387
Qy	2802	AACATGAGACCTGGCCCTCATGTTCAGAGACTCGGAAGAAATGCGGAGTTTGGGCTCT	2861
Db	3388	AACATGAGACCGACGTATGATGT-----TCGACAGAGGTCA	3423
Qy	2862	CTCAGCCAGGTTACTACAATGCAACAGTCTGTGATGTGCTCAGACAGGTGTCATATA	2921
Db	3424	CTGAGTCAAGGAACACAAATTAACATGCTGAGATG---TCAGGAGGTGCTCACAA	3480
Qy	2922	AAGGGGAGCTGCTAGTCTCTTCTTTCATATGCCAAAAAGCTTTATGAGATGCCCCAATG	2981
Db	3481	AAAAGGCAAGCCGACNCTCTGTGCTTAATGCCAAGACTATATGAGATGCCCAATG	3540
Qy	2982	GCTGGAAGATGAAGCAGTACCTTTTCCATTTTGGAGCTAGAAATGACGAGAGAGTCTT	3041
Db	3541	GCAAGGAAGGTAAGCAGTATCTTTTCCAGTCTGATGTAGAGACAGATGAGGAAGATTG	3600
Qy	3042	CAGACATATATCTGCAAGTGTGAGCCAGCAACACATTCCTCTAAGATCTTGCTGATC	3101
Db	3601	CAGATGATCTATTNCAATGGAGACCTGCATATGTGACTTGCACAAGATTTAATGTAG	3660
Qy	3102	AAAAAGCTGTCAA---ATCCGAGACTCTCCAGTAGCTCCAGGGCAGGGTCAACA	3157
Db	3661	AAAAGATAGCAAGATCATCTGAAATGTCTCCAGTCCATAGAGTGTGCTGGCCAAAC	3720
Qy	3158	GAAAGCTAGTCCCTGGCACAGGCCCCAGCAGCACCAACGACGACATTAATATCAACA	3217
Db	3721	AACATAAGCCACTTTGATCAACCC-----CACAGATTAAGCA	3755
Qy	3218	GGGACTACAGGTTCCCGCCGCTGTCCCTTATCTTCCAGGAAGAAAGTCCCGTAAGGA	3277
Db	3760	GGTGCTTCAAGTGGCAGGTGTATTTTGCACCCCATCAGAGAAAGGACAAACAAAGA	3819
Qy	3278	TCTCCCACTTTTGGCATTAACCTCCACAAGCTTTTAAAAAAATTTCTTTTCTTGA	3333
Db	3820	CCCTGCACTGAAATCAAGATTACTCTCAGAAAGTTTATGAGAACAA-----CTGA	3867
Qy	3338	AGAAAGAGTTTGGAGGTGCAAGAAACAGCGCTGAAGATACAAATPACAAATGCAATCTTC	3397
Db	3868	AGAAATTA-----GTGTAAAGAAAGCATACAGAAGACATATTTCTGTGGCGCATGC	3918
Qy	3398	GCAGCTTTCTTCTCTCTACTTCTCCACAGAGTTCTCAAGGAAGGCTATACTTTGGC	3455
Db	3919	TTTACATCTAGTCTCCTCGATCTCTCTCAAGGCTCCCGTCACAAAGGTTTACACTTAT	3978
Qy	3458	TCCAGTGTACTGTGATATTTTTCAGATTCTGGTCACAGGAATTTCTTCAGCATG	3518


```

Db 3979 TCCATCAGCTAAATCATGACAACTTGTGTGACTCCAGCCATAGTAGATTTCTTACGGTTC 4038
QY 3518 CAGTATGTACCAATTCCTCTTTTGCATCGACAGTCTACATCGACAGATGAGAGCGG 3577
Db 4039 CAGCATCGGACCAATGTTGTGTGTGACTCCTCATGTGTGAGCTTCAAGATGAACGGTG 4098
QY 3578 CCAGAGGCACTTGTGTGACAGCTGCTGGAACAAACCTAGGAGTGGCAGAGATGAGAGCGG 3637
Db 4099 TTCCTCTCAGGGCCCTGGCAGTCCCTGTAATCCATCGAGGCACTTGGAAACACAGAGACGC 4158
QY 3638 GACCATGATTGAACCTGATAGTAAAGCTTGGGGTCTTATGACCAACATGCCAGGCGG 3697
Db 4159 TTCAGGATAGAGATCATAGTCAACATGCGCCCTGGGTGGAGACATCTTGAAGCATCTCT 4218
QY 3698 AAGCTTATATGCTACAGCTACAGTATTTCTCTCCACAGACAGACAGATTTCCAGAGA 3757
Db 4219 AATCAAGTCTTAACTGTCTCA-----TCGTCTGTGACCAATGAAGAAATTTCTCAAGA 4272
QY 3758 TCAGGGGATCGCGCTCATCTTGTATGCTGCTGACAGTGGCCCTGGAGCTGACATCATG 3817
Db 4273 GCATATCATTT-----ATGAAGCAGCTGTACAGTGGTCTGTGGAAGTTGGAGCTCTGT 4323
QY 3818 CTCAGTGGCTCCCATGATATATATACAGACGATCCAGCACAGAGAGCTGGGA 3871
Db 4324 TTCACGACAGCTCCATGACAGCACTTCCAAAGCCTTCCAAACCAAAAGCTGGGA 4377

RESULT 8
AF070570 1803 bp mRNA linear PRI 05-AUG-1998
LOCUS AF070570 Homo sapiens clone 24473 mRNA sequence.
DEFINITION AF070570
ACCESSION AF070570.1 GI:3387939
VERSION FLL.CDNA.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1803)
AUTHORS Anderson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.
TITLE A double adaptor method for improved shotgun library construction
JOURNAL Anal. Biochem. 236 (1), 107-113 (1996)
MEDLINE 96207227
PUBMED 8619474
REFERENCE 2 (bases 1 to 1803)
AUTHORS Yu,W., Anderson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W.,
Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.
TITLE Large-scale concatenation cDNA sequencing
JOURNAL Genome Res. 7 (4), 353-358 (1997)
MEDLINE 97264341
PUBMED 9110174
REFERENCE 3 (bases 1 to 1803)
AUTHORS Yu,W. and Gibbs,R.A.
TITLE Direct Submision
JOURNAL Submitted (05-JUN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA
FEATURES
Source
1. 1803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="I.M.A.G.E. Consortium clone ID 24473"
/sex="female"
/tissue_type="brain"
/clone_1lb="1lb"
/dev_stage="infant"
/notes="similar to Homo sapiens mRNA for KIA0313 with
GenBank Accession Number AB002311"
BASE COUNT 545 a 323 c 346 g 589 t
ORIGIN
Query Match 24.1%; Score 1581.4; DB 9; Length 1803;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 1693; Conservative 0; Mismatches 1; Indels 81; Gaps 3;

```

```

QY 4875 CAAGATATTCACCTGCTGCTGACATTCATGCGCGAAGGATTTTCTTAATCTCCT 4934
Db 1 CAAGATATTCACCTGCTGCTGACATTCATGCGCGAAGGATTTTCTTAATCTCCT 60
QY 4935 TTTAGATTTCAATCCAGTCTAGCTATGATCATTTGGATTAATGAGAAAGCTAGCCA 4994
Db 61 TTTAGATTTCAATCCAGTCTAGCTATGATCATTTGGATTAATGAGAAAGCTAGCCA 120
QY 4995 TTGAATCTTGGGGCTTTAAACCCACCAAGAGAGCAAAAGAAAACAAATGAATCTTT 5054
Db 121 TTGAATCTTGGGGCTTTAAACCCACCAAGAGAGCAAAAGAAAACAAATGAATCTTT 180
QY 5055 GAGTACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5112
Db 181 GAGTACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 5113 TAAAGATTTTGTTCAGAGATTAATATATATCATTTAATGATTACAGTATTTTAA 5172
Db 241 TAAAGATTTTGTTCAGAGATTAATATATATCATTTAATGATTACAGTATTTTAA 300
QY 5173 CCTTAAGTAGGGTTGCGACGCTGCTTTTGAAGAAACCAATATGCCGACAGGCTGGC 5232
Db 301 CCTTAAGTAGGGTTGCGACGCTGCTTTTGAAGAAACCAATATATGCCGACAGGCTGGC 360
QY 5233 CACACCAAGAGAGGAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5292
Db 361 CACACCAAGAGAGGAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 5293 CACCCCGAAGTCCCTATCTGGAAGTATGAATGTAAATTAATACCAAGACACC 5352
Db 421 CACCCCGAAGTCCCTATCTGGAAGTATGAATGTAAATTAATACCAAGACACC 480
QY 5353 TCATCTGCTCTTCCCAAGTGTGGGGTCTTCTGTAACCTGTTTGCACATGGCCAG 5412
Db 481 TCATCTGCTCTTCCCAAGTGTGGGGTCTTCTGTAACCTGTTTGCACATGGCCAG 540
QY 5413 GAGGGAAGTACAGACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5472
Db 541 GAGGGAAGTACAGACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 5473 GCGGATGCTCCGCTGCATTCATGATGATGAGCAACCCCATTTTAAGTATATTTCT 5532
Db 601 GCGGATGCTCCGCTGCATTCATGATGATGAGCAACCCCATTTTAAGTATATTTCT 660
QY 5533 TCAATTTTGTATTTAAGAGTGTAGGTTTG-TTTTGTGTTTTTGTGTTTTTAA 5591
Db 661 TCAATTTTGTATTTAAGAGTGTAGGTTTG-TTTTGTGTTTTTGTGTTTTTAA 720
QY 5592 GAAACATTTATACGTGATAGCATTCGACAGGAAGAGCTGGAGTTCGACTAATGC 5651
Db 721 GAAACATTTATACGTGATAGCATTCGACAGGAAGAGCTGGAGTTCGACTAATGC 780
QY 5652 CAGCTGTTTATACGCTCTTCAAGACAGCTC----- 5684
Db 781 CAGCTGTTTATACGCTCTTCAAGACAGCTCCTTTATGAAATGGCATTTAGGAATA 840
QY 5685 -----CCTTATTTG 5693
Db 841 AACCAAGCTTTAAAGCTGATTAAGATCAAAAACCTGTTAGACATGCCAGCTTTATTTG 900
QY 5694 AATTCGATTTAGGAAATTAACAAGCCTTTTAACGTATTAAGATTAACAAGCCTGTTAG 5753
Db 901 AATTCGATTTAGGAAATTAACAAGCCTTTTAACGTATTAAGATTAACAAGCCTGTTAG 960
QY 5754 ACATGCGAGCTTTGAAGGACAGTATGACCAAGAGACTAACCTCAAGTGGCTTTAG 5813
Db 961 ACATGCGAGCTTTGAAGGACAGTATGACCAAGAGACTAACCTCAAGTGGCTTTAG 1020
QY 5814 GACGCTGATATAGAGAGGCTTAAGTATGACCAAGCATCTGCTACAGCTGCTATTAAAC 5873
Db 1021 GACGCTGATATATAGAGAGGCTTAAGTATGACCAAGCATCTGCTACAGCTGCTATTAAAC 1080

```


OY	465	GCTTTTGCCAAATATGACAAATGTAGTAGGCCGAACACTGTGTCTGTGATGTGGTTCGCA	524
Db	838	GCAATTGCAAAACAGACCATGTCTGTAGAGAGAACACTGTCTCAGATGATGATTTTTTGAA	897
OY	525	GTGGGGAAGAACCGAGGACCATAGTGTTAATAGATGTGAAGACGTGACCTCCGGTCA	584
Db	898	GTGTAGACAGCGCTGAGCTATTATTCTTGAAGATGGCCAAGACTTGACTCAGTGGAT	957
OY	585	GTCATCTCAATGAGATGTGGAAAGTGACTTATCCAGATGAGAAAAAGCAAAATACTGTGC	644
Db	958	GTTATTTTAAACGGCACTGTGGAAATCAGTCATCCAGATGAGAAAAGTTGAAAATTTGTGT	1017
OY	645	ATGGAAATAGTTTTGTGTCTCTCCTACCATSGACAAAGAATACATGTAAGAGAGTGTG	704
Db	1018	ATGGAAATAGTTTTGGAAATTACTCCACACTGTGATMACCATACATGCATGAATTGTCTC	1077
OY	705	AGAACAAAGGAGATGATCGCCAGTTGTGTGATGTGCCACCAAGATTACTGCCGTATT	764
Db	1078	AGGACTAAAGTAGATGTATCTACGATTTGTGTGATGTGCCACCAAGATTATTGGAAATT	1137
OY	765	CTCATCAAGTAGAAGAAAGCAATGCAAAAAAGTTGAGAGGAAGAGAGATTGTATGTGTG	824
Db	1138	TTTAAACCATGTGGAAAAAATTAACCCATTAAGTTTGGAGAGAGAGGAATTTGTATGGTA	1197
OY	825	AAGAAGACCGAGAACTTGATCGAACTGGAACAAGAAAGGACACATTTGTCATCAAGAGT	884
Db	1198	CATGACATCGGGAATAGACCGGAGTGGAACCAAGAAAGACATTTGTGATCAAGSCA	1257
OY	885	ACCTCAGAAAGGTTAACATGCAATTTGGTGGAAAGACCATTCAGTACGTATCCAACTTC	944
Db	1258	ACACCTGAGCGTCCATAAATGCAATTTAATAGAAACATTTCCATGCTGATCTCAACTTAT	1317
OY	945	ATTAGAAGCTTTCTGTGACCTGTATAGACCTTTCTTTAGCCCAAGTGAAGTGGGCAAA	1004
Db	1318	ATAGAAAGTTTTCTATTAACTTACAGSACATTTCTTGAAAGTCCTTTGGATTTGGGATC	1377
OY	1005	AAGTTATTGGAGTGTTPAATGACCCGAGCCTCAGGGATTAAGTTAACAGGGTACTATTTA	1064
Db	1378	AAACTATTGGAAATGTTTAAAGATCGACAGCTTAAGAGATTAAGGTGACACGGATTGATTA	1437
OY	1065	TTTGGGCTAATATATCACTTCATGACTTTTGAAGAGATCCTTCGACATGACTGATTTTTA	1124
Db	1438	TTTATGGGTAATTAATCAATTTTAATGATTTTGAAGGTGACCTTCGATGACTGATTTCTA	1497
OY	1125	GAAAGATTGGAACCATCTGGAAGAAGAGAAATGGTGGACACTTAAGGCTGTGATAT	1184
Db	1498	GAGGAATTTGAAAAAATCTGGAAGATACAAAGATTAATGTGTCATCTCCGGTTATTGAAAT	1557
OY	1185	ATCGGTGTGTCTGTAAAGCAAAAAGAGAATGTGATGAGTTTAACAAAMCATCCGAGAA	1244
Db	1558	ATTGCGTGTGCTGCAAAAGGCTAAGTGGACAGGTTGTCTGCAAAAAGGCTTCCGGCAG	1617
OY	1245	GCTCTTTGCTTTATCTTACTTGTGAAGCTCTGACAAAGGATTTGGAACTTTTGTGAC	1304
Db	1618	TCCCTCTACAAATTCACGCTTAATGGAGGAGTGAAGAAAGGATTTGTGATTTTGTGAA	1677
OY	1305	AGTGAAGTTAGGTAGTACCAAGACAGCAAGGAGGCTTAACACGGGGGATCAGATATTA	1364
Db	1678	GGAATGAACTGTGTACGGAAGCTGCTGATTCAGGACTAAACGTGTGTATAGATTAATG	1737
OY	1365	GAAGTAAATGGCCAAAACCTTTGAAAACATTCAGCTGTCAAAGCTATGGAATTTTATGA	1424
Db	1738	GAAGTAAATGSAACAAACCTTTGGAATATTAACATTTATGAAAGCGGTGAATTTTGGAG	1797
OY	1425	AATAACACACATTTATCTATCACTGTGAAACCAATTTATTGTATTTTAAAGAACTTCTA	1484
Db	1798	AATAATATCTATCTTGGACTTACGTGTAAGAACCMACATTTTGTGTCTCAAAGAGTTACT	1857
OY	1485	ACAAGATTGTAGAAAGAAAGAAATGTGGCCCCCACCTCTCTAAATTTGTGTACATTT	1544
Db	1858	TTTAG---GACTGAAACAAGAAATCTGGCGTTTCTCTCATATTTCCCAAAAT---TCCAAA	1911
OY	1545	AAAAAGGCACGTCTCATCCATTCAGATCTTGTGTAGATGTGAACAGGTGATPAGA	1604

Db	1912	AAAAAAGTATCCCATTTCTATCCAGCATGTGBCCGAGGAGATTTGAACAGCATCACAG	1971
Qy	1605	CTTGAAAAAGTGAACAAAAAAACTTAAAGCCAAACACTGTGSGAGGAAGAACAGCTGAA	1664
Db	1972	GAGAAAGGAGACTAAGAAAGTTTAAAGCAAAATACTGCTTCACGTGAGAAACAAATCAGG	2031
Qy	1665	AAGATACTCGACAAAGACTCGGATCGAGTATCTTGCCACAGAAACCAATACATGATATTGG	1724
Db	2032	AAGATTTTGGATTAACAGCATTTAGTATCTTGCCTCCAAGACTATTTAGTGTGAGGC	2091
Qy	1725	ATTGCTCAGTCTCAAGTATGACAGCATATGTATGATTAAGCGACAAAGCATCCCAACT	1784
Db	2092	CTAAGCCAATACAAAGATGACAGCAATGTGGGAACAAAGCACTGTATGAGCATGTCTGGCT	2151
Qy	1785	GCATTTGCGCTGACAGTGGAACTTATATCATCCAGTAATCCATGATTTTATGACATCATCAT	1844
Db	2152	ATTAATGCCATTCCTGGAAACACTCTATCCAGAGCCGTGATCTCTCGAGCTCCACCC	2211
Qy	1845	CGCATTTTACACTTCAGTGCCTACTCCGTGATTCGCCAGATCAAGTGTGAAGGTTTATAG	1904
Db	2212	AGTATGTGGATTTTTCATCTCTCTCAGATATCTCGATCAAGTTATTAAGAGTTTTCAAA	2271
Qy	1905	GCTGATGACAAAGCCGCTACATCATGATCAGTAAGAGACATPACAGCAAGAGAGTGTCTC	1964
Db	2272	GTGATGACAAAGTGTCTCATATTATCATCAGTAAGACACCAAGCTAAAGAACTAGTT	2331
Qy	1965	ATTCAAGCTATCAGGAGATTGCTGTTACTGCCACCCGGATCATATTTCACTATGTAG	2024
Db	2332	TTTCAATGCTGTTCAATGTAATTTGGTTTGACCGGATCCGACACATATTTCTCTGTGAA	2391
Qy	2025	GTCCTGTGCACACCTGAGGGAGTATCCAAACAAAGAAAGCACTTCAGATAGCTTTCCAA	2084
Db	2392	GTTTCTGTACTCTTGAGGGTGTCTATTAACAGAGAAAGCACTTCAGATAGCTTCTCCAAA	2451
Qy	2085	CTTGACAGACAAATACACTGAGTGGAAAGTATTTATCTGAAAAACAATGGAAACAGAA	2144
Db	2452	TTAGCTGATAGAAATTCACATCAATGGAAGTATTACTTAAAAAATPACATGGAAACAGAA	2511
Qy	2145	ACTCTTTGTTAAGTGAAGATGCTCAGAGATTTTGAAGAGAGTCAAAATTTCCCTCTT	2204
Db	2512	ACCTTATGTTTGAAGTGAAGATGTCTCAAGAACTGTTAAGAAAGCCAGCTATTCATGCTG	2571
Qy	2205	CAGCTCAGACACTGGAAGTGTGCAACACAGCTCTCATGCGAAATTTTAACTCTTGTGC	2264
Db	2572	CAGCTCAGATACCATTTGAGGTGGCCACCAGCTCTCAATGAGGACTTTGATTTGTTTGT	2631
Qy	2265	AACATTGAACCTACTGATATATAGATGATTTATTTAACTCAGATCAAAAACCACTGT	2324
Db	2632	AATATTTGAACCGACTGAGTACATGATGAGCTTTTAAGTAAATTCCAAACAGGAAT	2691
Qy	2335	GCCAACTCGAAGAGATTTGAAGAATCATTTAACCGAAACATTTTGGTACACTGTAA	2384
Db	2692	ACTATTTGAAGAGTTTTGAGGACATTTGTAAACCAAGAGACATTCGGGTTCCCTCAAAA	2751
Qy	2385	ATTTCACAGAAACAAACCGAGTGAAGAGATGAACATATTAAGCATTTCTCAAGATA	2444
Db	2752	ATTTTAACCTGAACAAATCAGCTCAACCGATGAABATTTATTAAGCATTTTATTAATTT	2811
Qy	2445	GCACCTGACTATAGGAATCGAAGAAATTTTAACTCATGTTTGGCATCATCATAGTGGCTTA	2504
Db	2812	GCACCTCATTTGTCAAGATATGAAGAACTTCAATTTTCATGTTTGSCAATATTAAGTGGCTTG	2871
Qy	2505	AACCTGGACCAAGTGGCAAGACTGCGAAGCACTGGGAGAAACATTCCAATTAATACAA	2564
Db	2872	AACCTGGCATCTGTAGCAAAACACTCAGAGAACTTGGGAAAAGTTTCCCAAGCAAAATACAG	2931
Qy	2565	AAACATTTCAAGATCTCCAAGACGTGTTGATCTCTCCAGAAACATGGCAAAATATGCT	2624
Db	2932	AAACATCTTCAAGATCTACAGAGCAATTTTGTATCTCTGTGAACATGTGGCAAAAGTATGAA	2991
Qy	2625	AATGTTTCATATCTCAAAATCTACAACTCCCATATCTCTATCTCCAGTATTCAAA	2684

Db 2992 AATATCTAGTAGTCAAGATATGCAAGCCTCCAAATTATTCCTCTGTTGTCACG 3051
QY 2685 AAGATCTCACTCTTCTTCGAGAGAAATGACTCAAAATTAACGGCGCTGCTCATATTT 2744
Db 3052 AAGATATGACATTTCTTACATGAAAGAAATGACTCCAAATAGATGGTTAGTAACTTT 3111
QY 2745 GAGAGCTAGAGATGATTGGAAAAAGAAATGTCACAGCTGGCCGCAATGGCTAGTGAAC 2804
Db 3112 GAGAGATTAGAAATGATTTCCAAAGAAATCCGCCAAGTTTTCGATGACTTTCGCTAAC 3171
QY 2805 ATGAGACCTGCCCCCTCATGTTTCAGAGACTCGAGAGAAAGAAATGGCGAGTTTGGGCTCTC 2864
Db 3172 ATGAGACCCACTATGATGTTT-----TCACAGAGAGTCTACTG 3207
QY 2865 AGCCAGGGTATGACAAATGCAACAGTCTTGATGTTGCTCAGACAGGTGGTCAATAAAG 2924
Db 3208 AGTCMAAGGAAGCACAAAATTCMAACATCTCTGATGT--TCAGGAGGTGCTCACAAAAA 3264
QY 2925 CGGGTAGTGTGTTGTTCTTTCGATGTCAGAAAGAGCTTTATGAAATGGCCCAAAATGCT 2984
Db 3265 AGGCGACCGCGCAGCTCTGCTTAAATGCAAGAGCTATATAGATGATGCCCAATATGCA 3324
QY 2985 CGAAAGTGAAGCAGTACTTTCCAATTTGAGCTAGAAATGACGAGAGAGTCTTCAG 3044
Db 3325 AGGAAGGTGAAGAGATATCTTTCCAGTCTCGATGATAGACAGATGAGAGAAAGTTCCAG 3384
QY 3045 ACATATCTCTGAGTGTGAGCCGACGACCAACACATTTGCTTAAGATCTCTGGTGAACAA 3104
Db 3385 ATGATGTCATTAACGATGGAGCGCTGATATGTTACCTTGACCAAGAAATTTAAGTGAACAA 3444
QY 3105 AAGCTTCTCAA---ATCCGAGACCTCTCCAGTACTCCAGAGGCGGCTCAACACAGAAA 3161
Db 3445 AGATCAGCAAAATGATCTGAAATGTCTCCAGTGTGATGAGGTGAGTGGCCAAACAACT 3504
QY 3162 GCTCAGTCCCTGCCACAGAGCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGGA 3221
Db 3505 AAGCCCACTTTGCAATCAACC-----CACAGAGTAAGCCAGGTG 3543
QY 3222 CTACAGGTTCCCGCGCTGCTTATCTTATCTTCACGAGAAAGATGGCCCTAAAGATCTC 3281
Db 3544 CTTCAGGTGCGCAGCTGTTAATTTGACACCCCTTCAGAGAAAGGAGCAACAAAGACCT 3603
QY 3282 CCACCTTTTGGCATMACTCTCCACAGCTTTAAAAAAATTTCTTCTGCTGAAGAA 3341
Db 3604 GCACGTGATACAGTTTACCTCAGAAAGTTTAAAGCAA-----CTGAGAA 3651
QY 3342 GGAAGTTTGGAGCTCAGAGAAACAGGCTGAAATATCAATATGATGATCTTCCAG 3401
Db 3652 ATAA-----GTGTAAGAGCAATACAGAGACACTATTTCTGTGCGCTCATCTTTA 3702
QY 3402 CTTCCTCTCTCTCTCTCTCTCTCCACAGAGTTCTCCAGGAAAGGCTTATCTTGGCTCC 3461
Db 3703 CATCTCTCTCTCTCTCTCTCTCTCCAGAGCTCCCTCACAAGGTTTACACACTTATCCA 3762
QY 3462 AGTGACTCTGAGTAATTTTTCAGATTTCTGTCACAGTGAATTTCTTCACATCCAGT 3521
Db 3763 TCAGCTAAATCTGAACTGTCTGACAGCCATGATGAGATTTCTTCAGCGCTCAGC 3822
QY 3522 ATTTGTTAGCAATTTGCTTTGACTCAGTGGCACTTCTACGACAGATGAGAGCGCCAG 3581
Db 3823 ATCTGTAGCAATTTGTTGTTGACTCCATGTTCTCAGCTCTACAGATGAACGGTGTTC 3882
QY 3582 AGGCAATTTGTCAGATGCTGGAAGAAACCTAGAGATGGCAGATGAGAGCGGAGCC 3641
Db 3883 TCTCAGAGCCCTGCGAGTCCCTGAATCCACTGGGCACTTTGAAAAACAGACAGAGCTTCA 3942
QY 3642 ATGATGAACCTGATCACTATAGCTTGGGTCCTATGACCAACATGTCGAGCGCGAGGC 3701
Db 3943 GGGAATAGAGATCATATGCAACATGAGCCCTGGGAGCACTCTTGAAGCCATCTCTAATC 4002
QY 3702 TTATATGCTACAGCTACAGTATTTCTTCTCACAAGCAGAGAACTTTCCAGATCAG 3761
Db 4003 AAGTGTTAGCTGTCTCA-----TCGTCTGTGAGCAATGAAGATTTTCTCAAGAGAT 4056

QY 3762 GGGAGTCGGCGTCACCTGATGCTGTGACAGTGGCGCGGAGCTGGACATGCTCA 3821
Db 4057 ATCAAT-----ATAGACAGACTGACAGTGGTGGGAGAGTTGACCTTCGTTCA 4107
QY 3822 AGTGCTCCCATGATATATACAGACATCAGCACCAAGAGCTGGGA 3871
Db 4108 AGCAGCTCCCATGACAACTCCAAACCTTCCAAACCCAAAAAGCTGGGA 4157

RESULT 10
AF478469 4176 bp mRNA linear PRI 25-FEB-2002
LOCUS Homo sapiens Rap1 guanine nucleotide-exchange factor PDZ-GEF2B
DEFINITION mRN, complete cds.
ACCESSION AF478469
VERSION AF478469.1 GI:18874699
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Kuiperij,H.B., de Rooij,J., Reijman,H., van Triest,M.,
Wiltinkhofer,A., Bos,J.L. and Zwartkruis,F.J.T.
TITLE Characterization of the PDZ-GEFs, a family of guanine nucleotide
exchange factors specific for Rap1 and Rap2

JOURNAL
REFERENCE 2 (bases 1 to 4176)
AUTHORS Kuiperij,H.B., Zwartkruis,F.J.T. and Bos,J.L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2002) Physiological Chemistry, University Medical
Centre, Universiteitsweg 100, Utrecht 3584 CG, The Netherlands

FEATURES
source
1..4176
/organism="Homo sapiens"
/db_xref="taxon:9606"

CDS
1..4176
/note="PDZ-GEF2"
/product="Rap1 guanine nucleotide-exchange factor
PDZ-GEF2B"
/protein_id="AA179916.1"
/db_xref="GI:18874700"

/translation="MNSPYDPGARQALRKKPEPTPEDLNTIYSLHGMETLSNLRH
OLRMSARARYEYSGNOYLPFCESTIARWYLLSGVLYGSMVLPSCFGKQFGK
RGCDCLVLEPSEMIYENAKNEDSLREIPAROSRRPRKINIKGROTTIDVEY
NYSLSLPADLTLMHLETPHPOVTHYSSQSCSTASQSGSSLSDIYOATSEPVGY
DLTRLEPGVDEDEDEDEDETRDPLQGRDLRECELEKREPADKTDIDQLSEFMH
QLPAFANMTMSVRELCSYVIEVEVQAGAILLEDGELDSWYLLNGTVEISHPDGK
VENLEWNGSEFGLPTLDKQYMGIVRTKWDQCFVCIADQWRLINHEVNTKRVK
EGEIVVHEHRELDGSGTRGKHIVIKAPERIMHLEHSLVDPYTEDFLVYRT
LESPLDVHGKILLEMFKISLRDKVTRVIRYIAWNNHNEDEGPATRLPEPEKLEID
TKNGHLRLNLNTACAKAKKAROVYLOKASRESPLOFSLNGSEKFGYFVCGE
AADSGLKRGDDIMEVNGENITETKAAVELIRNTHLALVYKNIYFKELPRTBO
EKSGVPHIKPIAKKSNRHSIQHVEGDIQETSQESKRVKNAKNSGGRNKRIRKTLDK
TFEESILPKLFEFDGSLQSDSIVITKQKSHSLAIMPPTLSSSSDLDLQPTSM
DESNPDIDPOVIRVEKVDQDSYIIISDITAKEVYHVAHEFLTASDPSYLSCEV
SYTPREGVIRKIDPOEKLADRIQNGYLYKNNMETETLCSDDAOELVEQSOLM
LOLSTIEVATOLSMRPFDELFRNIEPREYEDLFLKNSKGTGNHLKRPFDIVNOETFMV
ASFTLITEAQOLRMKLIKHTIRIALHCECKKPFNSMEFALISLNLASVARKLGYEKL
PSKYEKHLDQDIDFDSNRMAKRYNLISSQSMOPRILPLFVVRKDKMTFLHEGNDK
VGILVFERKILMIISKEIRGVVYMTSNMDDPAMFQORSLSQESTNSMLDVOGANK
BARRSKSLSPKLYEDQAKRKQVLYSLDVEDEKFOMLSLOMPAYGTLTNLS
EKRSKSSLSMSPVPMRSAGCTTKAHLHOBHRVSOYLOVPAVYVLLHPIRKKGOTKDPALN
TSLPOKVLGTTEIEISGKKTEDPTIYASLSHSPASPSPKHYGSLISHSKSTISG
QSCPGIGCATLQKTLIQITRTAKRTDTEFAETENRBDTSCENTTRKRMTSPFRLK
ERMLSRERLVNSQKEDTDHNOATECEKRVKGSNIKDEKSAIFNSQSGNSNLNC
FYTRFSKRRKRTL"

BASE COUNT 1378 a 776 c 944 g 1078 t

Query Match 22.9%; Score 1502.6; DB 9; Length 4176;

Best Local Similarity 68.7%; Pred. No. 0;
Matches 2158; Conservative 0; Mismatches 949; Indels 36; Gaps 5;

QY 45 AGCATATGTTTCTTCTATTTGAACCACTAGCATCCAGCTAACCATGAGATTAGGC 104
DB 418 AGAAGATTTTCGAAAAATTAATAAGAGAGCGCAACCACTACTATGATGTGGAG 477
QY 105 CAGCAGAGAAACACTCAGCTTCTGAGATTCACAAAACCTGATCTTGTGCACTGC 164
DB 478 GTTAAACAGCTATCTTCTCTTCCAGCTGATCTTACCAAGATGATCTCACAGAAACCCCT 537
QY 165 CACCCAGAGTGACCCAGCTTTCTTACAGCATTCAGATGTAGTATCACTAGTATTC 224
DB 538 CATCCAGAGTGACATGATGTCTTCTAGCACTGTGTTGATGATTCGCACTGACTCT 597
QY 225 GGGAGCAGAGCTTTTCTGATATCTACAGGCCACAGAAAGCGAGCTGTGATATGAC 284
DB 598 GGAAGCAGAGATTATCTGATATCTACAGGCTACGAGAGTGAGGTAGAGATGTAGAT 657
QY 285 CTGATGGGTTGGCAGAAACAGCAGTGTTCGGAAGCAGCAGCATGAGAGACATT 344
DB 658 TTGACACGCTTCTCAAGAGACCTGTTGATTCGAGGATGACGAAGAGATGAGAG 717
QY 345 GAGAGAGCATCAGATCCTCTGATGAGCAGGACATTGTGAGACTGCCTAGAGAGAC 404
DB 718 ATTGATCGAAACAGATCCATTCGAGGGCGAGATCTTTCGAGAATGTCTGAAAAAGAA 777
QY 405 CCAATTGACCGGACATGATGATGATGAAACAACCTTGGAAATTTATGACACCGTTCCCT 464
DB 778 CCGCAGACAAAACTGATGATGATGAACAATTCGTGAGTTTATGACACCGCTCCCT 837
QY 465 GCTTTTGGCAATATGACATGATGATGAGGCGAGACTGTGCTGATGATGGTTGCGCA 524
DB 838 GCATTTGCAACATGACATGTCTGTAAAGAGAGAACTCTGCTCAGTGTGATTTTGAA 897
QY 525 GTGTTGAAAAAGCAGGAGCAATAGTGTAAATGATGTGAAGAGCTGGACTCCTGCTCA 584
DB 898 GTGGTAGACAGAGCTGAGCTATTATTTCTGGAAGATGGCAAGAGCTGACATGATGTAT 957
QY 585 GTGATTCCTCAATGATCTGTGGAAGTACTTATCCAGATGGAAGAAACAGAAATCTGTGC 644
DB 958 GTTATTTTAAAGCGCAGCTGTGGAATCAGTCAATGAGTGAAGTGAAGTGAATTTGTTT 1017
QY 645 ATGGGAAATAGTTTGTGTCTCTCTACCATGAGCAAAAGATATCAATGAAGAGATGATG 704
DB 1018 ATGGGAATATGTTTGGATTAATCTCCACTCTGTGATAGCAGTACATGCCATGGAATTGTC 1077
QY 705 AGAACAAGGTGATGATGCGCAGTTTGTCTGCATAGACCCAGCAGATTAATCTGCCATAT 764
DB 1078 AGGACTAAAGTATGATGTGTAGTTGTCTGCATAGCCAGCAAGATTAATGGAACAATT 1137
QY 765 CTCGAATCAATGAAAGAAAGACATGCAAAAAGTTGAAGAGGAAGAGATGTTGATAGTG 824
DB 1138 TTAAACCATGTGGAATAAAATACCATTAAGTTGAGGAAGAGGAAGAAATTTGATAGTA 1197
QY 825 AAAGAACACCGAAGACTGTGCAACTGGAACAAGAAAGGACACATTTGATCAAGGGT 884
DB 1198 CATTGACATCGGAGACTAGACCGGAGTGAACCGAAGAGACATTTGATCAAGGCA 1257
QY 885 ACCTCAAAAAGTTAAACAATGCAATTTGTGGAAGAGCAATTCAGTATGATCAACATTC 944
DB 1258 ACACCTGAGCGTCTCAATATGCAATTAATGAAGAACATTCATCGTATCCAACTTAAT 1317
QY 945 ATTAGAGACTTCTGTGACCTATAGGACTTTCTTTAGCCCATGGAAGTGGGCAAA 1004
DB 1318 ATAGAAGATTTTCTATTAATCTACAGCAATTTCTGAAGATCTTGTGGATTC 1377
QY 1005 AAGTTATTTGAGTGTATTAATGACCCGAGCCTCAGGATTAAGTTACACGGGTAGTATTA 1064
DB 1378 AAACATTTTGGAATGTTTAAGATCGACAGCTTAAGAGATTAAGTTGACACCGATTTGATTA 1437
QY 1065 TTGTGGTAATATATCATTTCAATGACTTTGAAGAGATCTTCAATGATCATGATTTTAA 1124
DB 1124

DB 1438 TTATGGTAATATATCATTTTAAATGATTTTGAAGTGTACCCCTGATATGATGATTTCTA 1497
QY 1125 GAAGATTTGAAACATCTGTGAAAGAGAAATGGGTGAGACCTTAAGCCTGTTGAAT 1184
DB 1498 GAGGAATTTGAAAAAATCTGGAAGATACAAAGATTAATGCTATCTCCGGTTATGAAT 1557
QY 1185 ATCGCTGTGCTGCTTAAGCAAAAAGATGATGATGACGTTAAACCAACATCCGAGAA 1244
DB 1558 ATTGCTGTGCTCAAAAGGCTAAGTGAGACAGAGTGTGTGCTGCAAAAGGCTTCCCGAG 1617
QY 1245 GCTCTTTGCTTTTATCTTACTTGAGAGCTCTGAGAAAGGATTTGGAATCTTTGTGAC 1304
DB 1618 TCCCTCTCAATTTCAAGCCTTAATGAGGAGTGAAGAGGATTTGATTTTGTGGA 1677
QY 1305 AGTGTGATTCAGGTGCAAGCAACTGAAGCAGCTTGAACGGGGGATCAATATTA 1364
DB 1678 GGAATGAACCTGGTGAAGCAAGCTGCTGATTCAGCACTGAAGCTGTGATCAGATTATG 1737
QY 1365 GAAGTAAATGGCCAAACCTTGAACATTCAGCTGCAAAAGCTATGGAATTTCTAGA 1424
DB 1738 GAAGTAAATGACAAACCTTTGGAATATTACATTTATGAAGCCGTTGAATTTTGAGG 1797
QY 1425 AATAACACATTTATCTATCACTGTGAACCAATTTATTTGATTTAAGAACTTCTA 1484
DB 1798 AATAATATCTATCTGACCTTACTGTGAAGACCAACATTTTGTGTTCAAGATTAATCT 1857
QY 1485 ACAAGATTCAGAAAGAAAGAAATGTGTGCCCCCACCCTTCTTAATTTGTTGACATT 1544
DB 1858 TTTAG--GACTGAACAAGAAATCTGGGCTCCATATTTCCAAAT--TGCAAAA 1911
QY 1545 AAAAAGGCCAGTCTACTCTCATTTCCAGATCTTGTGTATGATGATGAGAGTGTATGGA 1604
DB 1912 AAAAAAGTATGCTCATTTCTATTCAGCATGTGCCAGAGATTTTGAACGACATCAAG 1971
QY 1605 CTTGAAAAAGTGAACAAAAAAGTAAAGCCACACTGTGGAGAGAAAGCAAGCTGAAA 1664
DB 1972 GAGAAAGGAAGTAAAGAAAGTTAAAGCAAAATACGTTCAGGTGAAGAAACAATACAG 2031
QY 1665 AAGATCTGACAAGACTGGATCAGTATCTTGTCCACAGAAACCATTCATGATATTGGG 1724
DB 2032 AAGATTTTGGATTAACACAGATTTAGTATCTTCCCTCAAAAGCTATTTAGTATGAGAGC 2091
QY 1725 ATGTGTCAGTCTCAAGATGACAGCATAGTATGATTAAGGACAGAAAGCAATCCCACT 1784
DB 2092 CTAAAGCAATACAAAGATGACAGATTTGTGGACAGACAGCTGTAGGCTATGCTGCT 2151
QY 1785 GCATTCCTGTGAGTGAAGCAACCTATATCATCAATCTGATTTATTTGACAGTCAATCAT 1844
DB 2152 ATATGCCCATCCCTGAAACACTCTCATTCAGCAGCCCTGATCTCCGAGCCTACCAACC 2211
QY 1845 CGCATTTTACACTTCAGTGTCTACTCTGACTTGTCCAGATATCAATGCTAAGGGTTTAA 1904
DB 2212 AGTATGTGGAATTTTCCAACTCTCATGATATCCCTGATCAATGTTAAAGATTTTCAAA 2271
QY 1905 GCTGATCAGCAAGCCGCTCATCATGATCAGTAAGACACTCAGCAAGGAAGGTGCT 1964
DB 2272 GTGATTCAGCAAGAGTGTCTATCATTAATCATATGATTAAGACACCCAGCTAAAGAGTAT 2331
QY 1965 ATTCAAGCTATVCAAGGAGTTTGTCTTACTGCAACCCCGGATCAATATTTCAATATGTAG 2024
DB 2332 TTTCATGCTGCTTCATGAATTTGTTGACCGGTGATCCGACACATATTTCTCTGTGAA 2391
QY 2025 GTCTCTGTACACCTGAGGAGATTAATCAACAAGAAAGTCTCAGATACGCTTTCCAA 2084
DB 2392 GTTTCGTATCTCTTCAAGGCTGTCAATAAACAAGAAAGCTTCAGATCAGTTCTCCAAA 2451
QY 2085 CTTGACAGACAGATTAACAGTGAAGTATTAATCTGAAAAACAAGAGAAACAGAA 2144
DB 2452 TTAGCTGATAGAAATTAACATCAATGGAAGTATTTACTTAATAAATAACATGAAACAGAA 2511
QY 2145 ACTCTTTTGTTCAGATGAAGAAGTGTCAAGAGTTGTGAGAGAGATCAAAATTTCCCTCTT 2204
DB 2512 ACCTTATGTTCAAGATGAAGATGCTCAAGAACTAGTTAAGAAAGCCAGTATCATGCTG 2571

OY	2205	CAGCTGACACTGTGGGAAGTTGCAAACACAGCTCTCATTGCGAATAATTTTTTGACCTCTTCGC	2264
Db	2572	CAGCTGACACTGTGGGAAGTTGCAGCCACCAGCTGTCAATTAAGAGGACCTTGGATTGTCTTCGT	2651
OY	2265	AACATTGAACCTACTGTAATATATATAGATGATTTATTTAACTCAGATCAAAAACCAAGCTGT	2324
Db	2692	AATATTGMAACCGACTGAGTACATCATGACCTTTTAAAGTTAAATTTCCAATACAGSAAT	2691
OY	2325	GCCAACTGGAAGACATTTGGAAGAAGTCAATTACCACGAAACATTTTGGTAGATCTCGAA	2384
Db	2692	ACTCATTTGGAAGAGTTTGGAGACATTGTAAACCAAGACATTTCTGGTTGCCCTCAGAA	2751
OY	2385	ATTCTCAGAGAAACCAACCCAGCTGAAGAGATGAAGATCATATTAGCATTTTCATCAACATA	2444
Db	2752	ATTTTAAGTGAAGCAATACACTCACAACGANTGAAGATTAATTAAGCATTTTATTAATAATT	2811
OY	2445	GCATGCACTGTAGGGAATGCAAGAATTTTAAGTCATGTTTGCATATCATCATGCGCCTA	2504
Db	2812	GCATCTCATTTGTGAGATGTAAACACTTCATTCATGTTTGCATTAATTAAGTGGCTTG	2871
OY	2505	AACCTGGACACAGTGGCAAGACTGGGAACGACGTGGGGAAGCAATCCCAATTAATATAGAA	2564
Db	2872	AACCTGGCATGTGACGAAGACTAGAGAACCTTGGGAAAAATTACCAAGCAATATAGAG	2931
OY	2565	AAACATATTTCAAGATCTTCACAAACCTGTTTATGCTTCACAGAAACATGGCAAAATATCGT	2624
Db	2932	AAACATCTTCAAGATCTTCACAAAGACATTTTATTCATCTATACAAACATGGCAAAAGTATGA	2921
OY	2625	AATGTTCTCAATAGTCAAATCTACAACTCCCATATCCCTCTATTTCCAGTTATCAA	2684
Db	2992	AATATTTCTTAGTACCAAGTATGAGAGCTCCCAATATTTCCACACTCTCCCTGTTGTCAAG	3051
OY	2685	AAGATCTCACCTCCCTCCACGAAGGAATATGCTCAAAAGTAGAGGGCTGGTCAATTTT	2744
Db	3052	AAAGATATGACATTTCTACATGAAAGGAATATCTCCAAAGTAGAGGTTTAGTAATCTT	3111
OY	2745	GAGAAGCTTAAGAGTATTGCAAAAAGAAATTTGTCACGTTGGCGGAATGGCTTCACTGANA	2804
Db	3112	GAGAAGTTAAGAGTATTTTCCAGGAANAATCCGCCAAGTTGTTGCAATGACTTCTGCTAAC	3171
OY	2805	ATGGAACCTGCCCTCATGTTTCAGACTCGGAAGAAATGGCGGAGTTTGGGCTCTC	2864
Db	3172	ATGGACCCACGATATGATGTT-----TCGACAGAAGGTCACCTG	3207
OY	2865	AGCCAGGGAGTACAAATGCAACAGCTGCTAGATGTTGCTCAGACAGGTGTCATTAAGA	2924
Db	3208	AGTCAAGGAGACAAATTTCAAACTGCTGATGT---TCAGGGAGGTGCTCACAAGAAAA	3264
OY	2925	CGGCTACGTCGTAGTCTCTTCTCAATGCCCAAAAAGCTTTATGAAGATGCCCAATGSGCT	2984
Db	3265	AGGGCAGCGCCGAGCTCTCTGCTTAATGCAAGAAGAGCTATATGAGAGATGCCCAATGSCA	3324
OY	2985	CGAANAAGTBAGACATACCTTTTCCAAATTTGGAGCTAGAGAAATGAGACGAGAGATCTTAG	3044
Db	3325	AGGAAGGTGTAACCAATATCTTTCACAGTCTCGATGTAGAGACAGATGAGAGAGATGCTTAG	3384
OY	3045	ACATTATCTCTGACGTGTGAGGCGACACAAACACATTTGAGTCCCTAAGAAATCTGATGACAAA	3104
Db	3385	ATGATGTCTTTACATGAGGAGGCTGCAATATGATACCTTACCAAGAAATTTTAAGTAGAGAA	3444
OY	3105	AAGCTGTGCA--ATCCGAGACCTCTCCAGTAGTACCTCCAAAGGCGAGGCTCACACAGAAA	3161
Db	3445	ACATCAGCCCAAAATCATCTGAAATGTCTCAGTGCCTATGAGGTCAAGTCCCAACAACT	3504
OY	3162	GCTCAGTCCCTCCACAGCCCCA	3184
Db	3505	AAAGCCCACTTGCAATCAACCCCA	3527

DEFINITION	SEQUENCE
Mus musculus chromosome UNK clone RP24-49JB24, WORKING DRAFT	10 unordered pieces.
ACCESSION	AC124358
VERSION	AC124358.2
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 181796)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	The sequence of Mus musculus clone
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 181796)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS	Parkway, St. Louis, MO 63108, USA
TITLE	3 (bases 1 to 181796)
JOURNAL	McPherson,J.D. and Waterston,R.H.
COMMENT	Direct Submission Submitted (04-JUL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA On Jul 4, 2002 this sequence version replaced gi:21426479.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BB0491B24

----- Summary Statistics -----
Sequencing vector: MJ3: 0%
Sequencing vector: plasmid: 100%
Chemistry: Dye-primer ET: 0% of reads
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: Version 0.990319
Consensus quality: 175933 bases at least Q40
Consensus quality: 176501 bases at least Q40
Consensus quality: 177034 bases at least Q20
Insert size: 168000; agarose-fp
Quality coverage: 10.97 in Q20 bases; agarose-fp
Quality coverage: 9.38 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
I
* 5780: 5779: contig of 5779 bp in length
* 5780: gap of unknown length
* 5880: 21352: contig of 15473 bp in length
* 21353: 21452: gap of unknown length
* 21453: 44568: contig of 23116 bp in length
* 44569: 44668: gap of unknown length
* 44669: 77089: contig of 32421 bp in length
* 77090: 77189: gap of unknown length
* 77190: 100595: contig of 23406 bp in length
* 100596: 100695: gap of unknown length
* 100696: 125457: contig of 24762 bp in length
* 125458: 125557: gap of unknown length
* 125558: 148442: contig of 22925 bp in length
* 148483: 148582: gap of unknown length
* 148583: 180953: contig of 32371 bp in length
* 180954: 181053: gap of unknown length
* 181054: 181631: contig of 568 bp in length
* 181632: 181721: gap of unknown length

```



```
* 181722 181796: contig of 75 bp in length.
FEATURES
    source             Location/Qualifiers
                        1..181796
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /chromosome="X"
                        /clone="RP24-491B24"
                        1..5779
misc_feature           /note="assembly_name:Contig24"
                        5880..21352
misc_feature           /note="assembly_name:Contig25"
                        21453..44568
misc_feature           /note="assembly_name:Contig26"
                        44669..77089
misc_feature           /note="assembly_name:Contig27"
                        77190..100595
misc_feature           /note="assembly_name:Contig28"
                        100696..125457
misc_feature           /note="assembly_name:Contig29"
                        125558..148482
misc_feature           /note="assembly_name:Contig30"
                        148583..180953
misc_feature           /note="assembly_name:Contig31"
                        181054..181621
misc_feature           /note="assembly_name:Contig17"
                        181722..181796
misc_feature           /note="assembly_name:Contig6"
                        181722..181796
BASE COUNT            52422 a 37730 c 37489 g 53249 t          906 others
ORIGIN
    Query_Match        20.5%: Score 1348.6; DB 2; Length 181796;
                        Best Local Similarity 86.4%; Pred. No. 5.9e-310;
                        Matches 1794; Conservative 0; Mismatches 194; Indels 89; Gaps 24;
    4536 GAAGATGAACAACTTCTGCTGTTTGGAGCAGACTTTCTGGAAGCAGAGCAGCCAC 4595
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117523 GAAGATGAACAACTGCTGCTGTTTGGAGCAGCCAGCC-TCCTGATCCAGTGAAGCCAC 117466
    4596 CTGAAGGAGAGCACAAGAAAGAGCTCTGAGCATTTGGAGCTTGGAACTCAGCTTCTGAG 4655
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117465 C-CAAAGGAGAGCACAAGAAAGAGCTCTGAGCATTTGGAGCTTGGAACTCAGCTTCTGAG 117408
    4656 GAGCGTGAACCAATTTGCTCTCTTCCCTGCTTAAAGAGCATTTGGAGCTTCTTCTCCG 4715
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117407 GATGGTGAACCAATTTGCTCTCTTCCCTGCTTAAAGAGCATTTGGAGCTTCTTCTCCG 117349
    4716 TTCTTCTCTTCCCTTTCGATGTGAATATCTGTGAAGAAATTCGCTGCACTTTTCAGA 4775
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117348 TTCTTCTCTTCCCTTTCGATGTGAATATCTGTGAAGAAATTCGCTGCACTTTTCAGA 117289
    4776 CTATTGCTGTAATGACAGTGCAGCAATCTTTCGAGCTCCCACTGTTGCTGCTGCCA 4835
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117288 C-TTGTGCTGTAATGACAGTGCAGCAATCTTTCGAGCTCCCACTGTTGCTGCTGCCA 117239
    4836 CATCAGACATGATCATTTCCAAATTCAGATCATCACAACAGATGATGATCATCTGCTG 4895
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117238 CGTACAGCATATCATTTCCAAATTCAGATGCTC---ACAAGGCAATTCCTGCTGCGG 117182
    4896 CACTTCTCAATGCTGTGAAGATTTTAACTTCTCTTTAGATTTCAATCCAGTCTCT 4955
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117181 CCTTCTCAATGCTGTGAAGATTTTAACTTCTCTTTAGATTTCAATCCAGTCTCT 117124
    4956 AGCACTGATCTCATTTGGGATATAGAGAAAGTATGAGCATTTGAGTCTTGGGCTTTTA 5015
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117123 AGCCCTCGATCTATTTGGAGTGAAGTATGAGTATGAGTATGAGTATGAGTATGAGT 117064
    5016 ACCCACAAGAGAGCAAGAAATCAATGAATCTTTAGTATGAGTATGAGTATGAGTATGAGT 5075
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117063 ACCGACCAAGAGAGCAAGAAATCAATGAATCTTTAGTATGAGTATGAGTATGAGTATGAGT 117005
    5076 GTTTACATGCTCTCTTTTAAAAAATGATGTTTAAAGATTTTGTTCAGAGATTA 5135
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 117004 GTTTACATGCTCTCTTTTAAAAAATGATGTTTAAAGATTTTGTTCAGAGATTA 116951
```

```
    5136 ATA-TATATCCATTTAATGATTAAGTATTAAGTATTTTAAACCTTAAGAGGTTGCCACCT 5194
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116950 ATGTTATATTCATTTTATGATTAACATATTTATTTTAACTTAAGTATGATTTGCCACCT 116891
    5195 -GGTTTCTGAAAACCAATATATGCGGACAGAGGTGTGGCCACCAAGAAAGAGCGGAGAGA 5253
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116890 GGTTTCTGAAAACCAATATATGCGGACAGAGGTGTGGCCACCA-----GGGAGAGA 116837
    5254 CTTGCTTGTGACCTGTGCTTCCATGTCCTTGTGCTTCACCCGCGAAGTGGCTTATCC 5313
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116836 CTTGCTTGTGACCTGTGCTTCCATGTCCTTGTGCTTCACCCGCGAAGTGGCTTATCC 116777
    5314 TGGATGATGAATGTTAGCATTT-----AATACCAAGACACTTATCTGCTCTTCC 5368
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116776 CAAAGCATCAAGCTTATGCAATTTAAATTAACCAAGCTTATCTGCTCTTCC 116717
    5369 CAGTGTGAGGCTTCTTCTGTAATGATTTTGCATATGCGCAGGAGGAGGAGGAGTATGAGC 5428
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116716 CAGTGTGAGGCTTCTTATATGTAATGATTTTGCATATGCGCAGGAGGAGGAGTATGAGC 116657
    5429 CTTGTCTCTGTCGAGCCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5488
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116656 TGTGTCTCTGTCGAGCCTTAT-TAGGAGAGATGAGGAGGAGGAGGAGGAGGAGGAG 116598
    5489 CCATTTGATGATGAGGCAACCCCATTTTAACTATATTTCTTGTATTTTCTTAAATT 5548
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116597 CAATTTGATGATGAGGCAACCCCATTTTAACTATATTTGTTGTTTGTGTTTGT 116538
    5549 AGAGGTGAGGCTTTGTTTGTGTTT-----5575
    Db 116537 GGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 116478
    5576 -TTGTTTTTTTAAAGAGAAACATTTATACGTGATGATGATGATGATGATGATGATGATG 5633
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116477 TGTTTTATTTTAAAGAAACATTTATACGTGATGATGATGATGATGATGATGATG 116419
    5634 GATTTGAGGATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5693
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116418 GGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 116359
    5694 AATTGCTATGAGGATTAACCAAGCCTTTAAAGCTGATTAAGAAATCAAAACCTGCTTAG 5753
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116358 GATTGGACACAGAGAT---ATGCTTTAATCATATTTAAGTGTACCAACATGTTTAA 116303
    5754 ACATGCGAGCTTTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5813
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116302 ATACGCGAGCCTTTGCAA-AATGTTAGTACCAAAAGGAGGAGGAGGAGGAGGAGGAG 116244
    5814 GAGCGTGCATATGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5873
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116243 GAGCGTGCATATGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 116185
    5874 CTATTAATGAGCTG-AAATGAGCCCTCCACTCTATTTTGTGTTG-TTTTGGACAGACTCG 5931
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116184 CTGAATGATGATGAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 116125
    5932 GAAAGTGAAGGCTGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5991
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116124 GAAAGTGAAGGCTGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 116066
    5992 TTTTTCCTATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6051
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116065 -TTTTCCTATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 116008
    6052 ATTTTAAAGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6111
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 116007 ATTTTAAAGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 115948
    6112 AATTAATTTCTTGTATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6171
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
    Db 115947 AATTAATTTCTTGTATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 115888
```


Db 601 TTTAATGATTTTGAAGTGACCCCTGCTATGACTTCATTTCTAGAGAAATTTGAAAAAT 660
Oy 1143 CTGGAAGAGAGAAAATGGGTGACACCTTAAGCCTGTGAATATGCCGTGTCTGTA 1202
Db 661 CTGGAAGATACAAAGATGATGTCTCCGTTATTTGAATATGCGTGT-CTGCAAG 719
Oy 1203 GCAAAAAGAGATTTGATGCTTAACAAACCATCCGAGAAAGCTCTTGGCTTTATC 1262
Db 720 GCTAATGAGACAGGTGCTGCT--GCAAAAAGCTTCCCGAGTCCCTCACAATTCAG 777
Oy 1263 TTACTTGAAGCCTCTGAGAAGGATTTGGAATCTTTGTGACAGTGTAGATTGAGTAC 1322
Db 778 CTTAATGAGAGAGTGAAGGATTTGGTATTTTGTGAAGAGTAAACCTGTAGAC 837
Oy 1323 AAAGCACTGAAGCAGCAGCTTGAACCGGGGATCAGATATTAGAACTAAATGGCCAAAC 1382
Db 838 AAAGCGCTGATTCAGAGCTGAACAGTGTGATCAGATTAATGGAATTAATGACAAAAC 897
Oy 1383 TTTGAATACATTCAGCTGCAAAAGCTATGAAATCTTAGAAATTAACACATTTATCT 1442
Db 898 TTTGAGAAATTTTACATTTATGAAAGCCGTTGAATTTTGAGAAATTAATCTCATTTGCA 957
Oy 1443 ATCACTGTGAACCAATTTATTTATTTAAAGAACTTCTAAACAAGTTGTGAGAAG 1502
Db 958 CTCAGCTGTGAAGCAACATTTTGTGTTCAAAAGATTAATTTTGTAG--GACTGAACAA 1014
Oy 1503 AAAAGAAATGTCCT 1562
Db 1015 GAGAAATGTCCT 1071
Oy 1563 TCCATTCAGAACT 1622
Db 1072 TCTATCAGACATTCGTCAGAGATATTGAACAGGATCAGAGAAAGAAAGTAAGAAA 1131
Oy 1623 AAAAGTAAAGCAACACTGTGGAGAGAGAACAGCTGAAAAAGTACTCGACAAGCT 1682
Db 1132 GTTAAAGCAAAATCTGTTTCTGAGTGAAGAAACAAATCAGAAAGTTTGTGATTAACA 1191
Oy 1683 CGGATAGATATCTGTCAGAGAAACATCAATGATTTGGGATTTGGTCACTGCTCAAGAT 1742
Db 1192 CGATTTAGATCTCTGCTCAAGAACTATTAGATGAGGCTTAAGCCATACAAAGAT 1251
Oy 1743 GACAGCATGATGATTAAGGAGACAAAGACATCCCACTGATTCCTGCTGCTGCTGCTG 1802
Db 1252 GACAGCATGATGATTAAGGAGACAAAGACATGCTGCTGCTGCTGCTGCTGCTGCTG 1311
Oy 1803 ACCTTATCATTCAGTATCTGATTTATGACAGTCAATCATGCAATTTTGAAGCTTCACT 1862
Db 1312 ACATCTCATCCAGAGCCCTGATCTCTGACGCTACCCAGTATGTTGATTTTTC 1371
Oy 1863 GCTACTCTGACTTGGCAATCAAGTCTTAAGGCTTTTAAAGCTGATCAGCAAGCCGC 1922
Db 1372 AATCCTTCAAGATTCCTGATCAAGTTATTAAGGTTTCAAAAGTGAATCAGCAAAATGG 1431
Oy 1923 TACATCATGATCTGATGAGCACTATACAGCAAAAGAGTGTCTATGAGCTATCAGAGAG 1982
Db 1432 TACATCATGATCTGATGAGCACTATACAGCAAAAGAGTGTCTATGAGCTATCAGAGAG 1491
Oy 1983 TTTGCTTACTGCAACCCCGATCAATATTCATATGAGAGTCTCTGTCAACCTGAG 2042
Db 1492 TTTGCTTACTGCAACCCCGATCAATATTCATATGAGAGTGTCTCTGTACTCTCTGAG 1551
Oy 2043 GAGTATCAAAAGAGAGCTTCCAGATCAGCTTTCCAACTTGCAGACAGAAATACAA 2102
Db 1552 GAGTATCAAAAGAGAGAGCTTCCAGATCAGCTTTCCAACTTGCAGTATTAATTAACAA 1611
Oy 2103 CTGAGTGAAGATTTATCTGAATAACAAACAGAGAACTCTTTGTTAGATGAA 2162
Db 1612 CTGAGTGAAGATTTATCTTAATAAATTAACATGAAAGAAAGCTTATGTTAGATGAA 1671
Oy 2163 GATGCTCAGAGTGTGAGAGAGTCAAAATTTCCCTCCCTTCAAGCTCAGCACTGTGAA 2222
Db 1672 GATGCTCAGAGTGTGAGAGAGTCAAAATTTCCATGCTGAGCTCAAGTACATTTGAG 1731

Oy 2223 GTTGCACACAGCTCTCTATGGAATTTTGAACCTTTGCAACATGACCTACTGAA 2282
Db 1732 GTGGCCACCAGCTGATGAGGAGCTTTGATTTTGTGATATATGACCGACTGAG 1791
Oy 2283 TATATGATGATTTATTTAACTCAGATCAAAACAGCTGTGCCAAGTGAAGATTT 2342
Db 1792 TACATGATGACCTTTTAAAGTTAAATTCGAAACAGAAATTCATTTGAGAGCTTT 1851
Oy 2343 GAAGAGTATTAACAGAGAACTTTGGTGATCTGAAATTTCTGAGAGAAAC 2402
Db 1852 GAGGACATTTGAAACCAAGACATTCGCTGCTCAGAAATTTTAACTGAAGCAAT 1911
Oy 2403 CAGCTGAAGAGTGAAGATCAATTTATCATGAAATTAACACTGACCTGAGGAA 2462
Db 1912 CAGCTGAAGAGTGAAGATTTATTAAGCATTTTATTAATTAATTAATTAATTAATTA 1971
Oy 2463 TGCAGAAATTTTAACTCAATGTTTGAATCATGAGGCTTAACCTGGCAGCAGTGA 2522
Db 1972 TGTAAAGTCTCAATTCATGTTTGCATATATATAGTGGCTTGAACCTGGCATGTAGCA 2031
Oy 2523 AGACTGGAACGACCTGGAGAACTTCCAAATTAATAGAAAACTATTTCAAGATCTC 2582
Db 2032 AGACTGGAAGAACTTGGGAAAGTTACCAAGCAATATAGCAAAACATCTTCAAGATCTA 2091
Oy 2583 CAAGACTGTTTATCTCTCCAGAAACATGAGCAAAATATCTGATATGTTCTCATAGTCA 2642
Db 2092 CAAGCAATTTTATCTCATCTATGAAACATGGAAGATTAAGAAATATCTTATGATCA 2151
Oy 2643 AATCTCAACCTTCCATTAATCCCTATTCCTATTCAGTATTCAGAAAGATCTCCTCT 2702
Db 2152 AGTATCGGCTCCATTTATTCACCTTCCCTGTTGTCAAGAAATATGACATTTCTA 2211
Oy 2703 CACGAAGAAATGCTCAAAAGTATGACGGCTGCTCAATTTTGAAGAGTAAAGATAT 2762
Db 2212 CATGAAGAAATGCTCAAAAGTATGATGTTATGAATTTTGAAGATTAAGAAATGAT 2271
Oy 2763 GCAAAAGAAATGCTCAGCTGGCGAATGGCTTCAAGTGAACATGAGCCCTGCTCAT 2822
Db 2272 TCCAGAAATTCGCGAAGTGTTCGAAATGATCTGCTCAATGAGCCAGCTATGATG 2331
Oy 2823 TTTGAGCTGGAAGAAATGCGGAGTTTGGGCTCTCTCAAGGAGTATGACAAAT 2882
Db 2332 TTTGAGAGGAGAAAGAGTGTGCGGAGTGTGGGCTCACTGATGATCAAGAACACAAAT 2391
Oy 2883 GCAAGCTGATGATGCTGACAGAGTGTGATTAAGAAAGCGGATGCTGATGCTC 2942
Db 2392 TCAACATGCTGATGCT---TCAGGAGGCTGTCACAAAAAAGGAGCGCCAGCTCT 2448
Oy 2943 TTTCTCAATGCCAAAAAGCTTTTGAAGATGCCCCAATGGCTGGAAGTGAAGCAGTAC 3002
Db 2449 CTGCTTAATGCCAAAGCTATATGAGGATGCCCAATGCGAAGGAGGATGAGCTAT 2508
Oy 3003 CTTTCAATTTGAGCTGAAGATGAGAGAGAGTCTTGAAGATTTATCTGAGTGT 3062
Db 2509 CTTTCCAGTCTCATGTAGAGACAGATGAGAGAAAGTTCCAGATGATGATTTACAGTGG 2568
Oy 3063 GAGCAGCAACCAACATTTGCTTGAAGATCTGATGACAAAGAGCTGTCAAA--ATCC 3119
Db 2569 GAGCAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2628
Oy 3120 GAGACCTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3179
Db 2629 GAAATGCTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2688
Oy 3180 CCCC 3184
Db 2689 CCCC 2693

RESULT 13
AC118835/C
LOCUS AC118835 185108 bp DNA linear HTG 18-JUL-2002

DEFINITION Rattus norvegicus clone CH230-262L23, *** SEQUENCING IN PROGRESS
***, 63 unordered pieces.
AC118835
VERSION AC118835.4 GI:21747243
KEYWORDS HTG: HTGS.PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 185108)
REFERENCE 1
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barberia,J., Benton,J., Bimge,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burke,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Meheswari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Messey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenwo,S., Oguh,M., Okwuon,G.,
Orangunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peeters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swalek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tsanai,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlarczyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 185108)
Worley,K.C.
Direct Submission
Submitted (21-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 185108)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20330712.

Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GVTy

Center clone name: CH230-262L23

Summary Statistics

Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap, version 0.990229
Consensus quality: 127499 bases at least Q40
Consensus quality: 136160 bases at least Q30
Consensus quality: 141175 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1022: contig of 1022 bp in length
* 1122: gap of unknown length
* 1123: contig of 1015 bp in length
* 2138: gap of unknown length
* 2237: gap of unknown length
* 2238: contig of 1186 bp in length
* 3423: gap of unknown length
* 3524: contig of 1257 bp in length
* 4781: gap of unknown length
* 4881: contig of 1190 bp in length
* 6071: gap of unknown length
* 6171: gap of unknown length
* 7207: gap of unknown length
* 7306: contig of 1036 bp in length
* 7307: gap of unknown length
* 8639: contig of 1333 bp in length
* 8640: gap of unknown length
* 8739: gap of unknown length
* 8740: contig of 1500 bp in length
* 10239: contig of 1500 bp in length
* 10240: gap of unknown length
* 10339: gap of unknown length
* 10340: contig of 1257 bp in length
* 11596: gap of unknown length
* 11597: gap of unknown length
* 11696: gap of unknown length
* 13560: contig of 1864 bp in length
* 13561: gap of unknown length
* 13660: gap of unknown length
* 14957: contig of 1297 bp in length
* 14958: gap of unknown length
* 15057: gap of unknown length
* 15058: gap of unknown length
* 16246: contig of 1189 bp in length
* 16247: gap of unknown length
* 16347: gap of unknown length
* 16347: contig of 1827 bp in length
* 18173: gap of unknown length
* 18174: gap of unknown length
* 18273: gap of unknown length
* 19773: contig of 1500 bp in length
* 19774: gap of unknown length
* 19873: gap of unknown length
* 20937: contig of 1064 bp in length
* 20938: gap of unknown length
* 21037: gap of unknown length
* 22694: gap of unknown length
* 22694: contig of 1657 bp in length
* 22695: gap of unknown length
* 22795: gap of unknown length
* 24383: contig of 1589 bp in length
* 24383: gap of unknown length
* 24384: gap of unknown length
* 24484: contig of 1221 bp in length
* 25704: gap of unknown length
* 25705: gap of unknown length
* 25804: gap of unknown length
* 28163: contig of 2353 bp in length
* 28164: gap of unknown length
* 28263: gap of unknown length
* 29401: contig of 1138 bp in length
* 29402: gap of unknown length
* 29501: gap of unknown length
* 29502: contig of 2133 bp in length
* 31633: gap of unknown length
* 31634: gap of unknown length
* 31733: gap of unknown length
* 31734: gap of unknown length
* 33921: contig of 2188 bp in length
* 33922: gap of unknown length
* 34021: gap of unknown length
* 35757: contig of 1736 bp in length
* 35758: gap of unknown length
* 35857: gap of unknown length
* 37514: contig of 1657 bp in length
* 37515: gap of unknown length
* 37614: gap of unknown length
* 37615: gap of unknown length
* 39730: contig of 2116 bp in length
* 39730: gap of unknown length
* 39830: gap of unknown length
* 39831: gap of unknown length
* 41739: contig of 1909 bp in length
* 41740: gap of unknown length
* 41839: gap of unknown length
* 43994: contig of 2155 bp in length
* 44094: gap of unknown length

Db 192450 CCACGCTGCGGTATCATCTCCAAATTCACAGATGT---TACAGGTGATTTTCCCTGG 192394
 QY 4893 CTGCACCTTCATGCGCTGGAGAGATTTTAAATCTTCCTTTAGATTTTCAATCCAGT 4952
 Db 192393 GGACACCTTCATGCGCTGGAGAG---GTTTCCAAATCTTCCTCGGATTCACACAGC 192336
 QY 4953 CCTAGCATTGATCTGATGGGATATGAGAAAGCTAGACATGGAATCTTGGGGCCCT 5012
 Db 192335 CCTAGCCTCGATCTCATTTTGGAGATGAGAAAGCTAGCCTTTTGAATGCAAGG----- 192281
 QY 5013 TTAAACCCACCAAGAGAACAAAGAAACATGAAATCCTTTGATGACAGTGGCTTGCA 5072
 Db 192280 -----AAGACCGACAGATATGAAATCCTTAGAGTA--TGTCCTTGCA 192238
 QY 5073 CTGTGTTACATGCTCCTTTTAAAAAATAATGAGTTTAAAGTTTGTTCAGAGAG 5132
 Db 192237 CGGTGTTACAAACATCT-----TTAAATTAAGCTTAAAGATGTGTTGTCAGAGAG 192188
 QY 5133 TAAAT-ATATATCATTTAATGATATACAGT-----ATATTTTAAACCTTAAAGAGGTT 5186
 Db 192187 TAAATGTTACATCTATTTAGTGTATTTATTTAATGATCATTTTGAACCTTTAAGTATG 192128
 QY 5187 GCCAGCCTG-GTTTCTGAAAAACCAATATGCCGACAGGAGTGCCACACCAAGAGAG 5245
 Db 192127 GCCAGCCTGTTCTGAAAAACCAATATGCCGAGAGGTTTGTGTCGTCACACA--- 192070
 QY 5246 CCGGAGACCTGGCTGTGACCTGGCTTCCCATGTCCTTCTGTGTCACCCCGCAAGTG 5305
 Db 192069 ---GAAGACCTGGGACGATGTTCTGTCTCCCATGTCCTTGTGCTTGGCAATGAGTG 192013
 QY 5306 CCTATCTCGAGATATGAAATGTAGACCAATTA-----ATACCAAGACACTCATCTGC 5360
 Db 192012 CCTATCTCGAGATATGAAATGTAGACCAATTTAGTATACCAAGACTTCTCACCTGC 191953
 QY 5361 TCCCTTCCCAAGTGGATGGGTTCTCTGTAAACCTTTTGCACATGCGCAGGGAGGAA 5420
 Db 191952 TCCCTTCCCAAGTGGATGGGCTTAATGTAAAC---TACAGTGGCCAGGGAGGAA 191897
 QY 5421 CTAGGACCTGTGTCTGTCTGTGACCTTATGAGAGCAGAGACGTCATTTGGCGATGT 5480
 Db 191896 CTAGGACCTGTGTCTGTCTGTGACCTGCT-GAGGACGATGTGTGTGTGAGAACCT 191838
 QY 5481 GTCTCTCTCATTTGAGATGATGATGAGCAAAACCCATTTTAAAG---TTATATTTCTTGAT 5537
 Db 191837 GTCTCTCTCATTTGAGATGATGATGAGCAAAACCCCTGTCTCTAGATTTTGTGTTGAGT 191778
 QY 5538 TTTGTTAAATTTAGAGGTAG-----GTTTGTGTTTGTGTTTGTGTTTGTGTTT 5585
 Db 191777 TTTGTTAAATTTAGAGTTTAAAGCTTTTGTGTTTGTGTTTGTGTTTGTGTTT 191718
 QY 5586 TTAGAGAAACATTTAACTGATAGCATGATGAGTGAAGACGCTTGGATTTGGAGC 5645
 Db 191717 TATTTTAAAGAAACATTAAGTATGATGATGAGTGAAGACGCTTGGATTTGGAGT 191659
 QY 5646 TATGCGCAGCTGTT-TATACTGCTTTTCAAGACAGCTCCCTTTATGATTTGGCATTA 5704
 Db 191658 GATGCGCAGCTATTTTCGATGTTTCTTCAAGACAGCTCTCTTGTGATTTGGCACCA 191599
 QY 5705 GGGAAATTAACACCTTTTAAACGTGATTAAGATCAAAACCTGTTAGACATGCCAGCC 5764
 Db 191598 GAGAGT-----ATGCTTTTAAACATGAT-AGAGGTCAACAAACATGTTACATGACATAGCC 191544
 QY 5765 TTGTCAGACGATTTAGTACACCAAGACATTAACCTCAAGTGGCTTTATGAGCCTGCATA 5824
 Db 191543 TTGTCAGAACG-TGGTTAGTACCAAGACGATGTTAGAGGCTTTGTGATCTCTCACA 191485
 QY 5825 TAGAGAGGCGTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5884
 Db 191484 TAGAGAGAAACCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191425
 QY 5885 G-AAATGACCCCTCATCTTATTTTGTGTT-GTTTGCACAGACTCCGGAAGAGTGAAG 5942
 Db 191424 GAAATGACCCCTCATCTTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 191365

QY 5943 GCTGCCAATCTGAGTAGTACTCAATGTGAGAGACTGCTGCTGTGGATTTTTCAT 6002
 Db 191364 GCTGCCAATCTGAGTAGTACTCAATGTGAGAGACTGCTGCTGTGGATTTTTCAT 191308
 QY 6003 TAAATTCAGCTGATCATTTATTTAGTATGATGATGATGATGATGATGATGATGATGAT 6062
 Db 191307 TAAATTCAGCTGATCATTTATTTAGTATGATGATGATGATGATGATGATGATGATGAT 191248
 QY 6063 GGAATTCAGCTGATTTTTCAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 6122
 Db 191247 CAATATTCAGCTG-TTTCCTACCTGCTCCAAACATGACAGCTTATTTAATATCTCTG 191189
 QY 6123 TCTGTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6182
 Db 191188 TCTGTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191129
 QY 6183 ACATGTAATATGAT-----TATTTGACAGTTTATTTATATGATGATGATGATGATGAT 6228
 Db 191128 ACATGTAATATGAT-----TATTTGACAGTTTATTTATATGATGATGATGATGATGAT 191069
 QY 6229 CCTATGTCATATTAAGAAACAGAAATCTAGCTTACACAAAGTTGACAAATGTTATC 6288
 Db 191068 CCTATGTCATATTAAGAAACAGAAATCTAGCTTACACAAATTTTACAGGTATATC 191009
 QY 6289 TAAGCATTAAGTATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 6348
 Db 191008 TAAGCATTAAGTATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 190949
 QY 6349 GTGCAATATTAAGTATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 6408
 Db 190948 GTGCAATATTAAGTATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 190890
 QY 6409 TGTCTTTTAAAGAAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 6468
 Db 190889 TATCTTTTAAAGAAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 190831
 QY 6469 GTACATA 6475
 Db 190830 GTACATA 190824

RESULT 15
 AC023167
 LOCUS
 DEFINITION
 AC023167
 AC023167.9 GI:16118059
 VERSION
 KEYWORDS
 HTG; HTGS; PHASBL.
 SOURCE
 ORGANISM
 Mus musculus
 Mus musculus

REFERENCE
 AUTHORS
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 148466)
 Melzer, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
 Dederich, D., Thomas, S., Okwono, G., Carllock, C., Garner, T.,
 Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
 Bunay, C., Bunac, C., Burkett, C., Chacko, J., Chen, Z.,
 Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
 Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
 Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,
 Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
 Kovar, C., Liu, J., Liu, W., Louisegeed, H., Lozano, R.J., Martin, R.,
 Massey, E., McLeod, M.P., Mel, G., Moore, S., Morgan, M., Morris, S.,
 Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogden, M., Parish, B.,
 Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
 Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
 Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
 Worley, K. and Gibbs, R.
 Direct Submission
 TITLE
 JOURNAL
 REFERENCE
 2 (bases 1 to 148466)

AUTHORS
TITLE
JOURNAL

COMMENT

Worley, K.C.
Direct Submission
Submitted (09-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 14, 2001 this sequence version replaced g1:11079356.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: MAM

Center clone name: RP23-361K18

Summary Statistics

Sequencing vector: M13; L08621
Chemistry: Dye-primer Bodipy: 68% of reads
Chemistry: Dye-terminator Big Dye: 32% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 139795 bases at least Q40
Consensus quality: 171357 bases at least Q30
Estimated insert size: 165623; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-IP estimation
Quality coverage: 2.4x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length.

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 36 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

```

1 10583: contig of 10583 bp in length
* 10584 10683: gap of unknown length
* 10684 20426: contig of 9743 bp in length
* 20427 20526: gap of unknown length
* 20527 29062: contig of 8536 bp in length
* 29063 29162: gap of unknown length
* 29163 36210: contig of 7048 bp in length
* 36211 36310: gap of unknown length
* 36311 42380: contig of 6070 bp in length
* 42381 42480: gap of unknown length
* 42481 47818: contig of 5338 bp in length
* 47819 47918: gap of unknown length
* 47919 53861: contig of 5943 bp in length
* 53862 53961: gap of unknown length
* 53962 58028: contig of 4067 bp in length
* 58029 58128: gap of unknown length
* 58129 61169: contig of 3041 bp in length
* 61170 61269: gap of unknown length
* 61270 65935: contig of 4666 bp in length
* 65936 66035: gap of unknown length
* 66036 68826: contig of 2791 bp in length
* 68827 68926: gap of unknown length
* 68927 71855: contig of 2929 bp in length
* 71856 71956: gap of unknown length
* 71957 76368: contig of 4413 bp in length
* 76369 76468: gap of unknown length
* 76469 79877: contig of 3409 bp in length
* 79878 83981: contig of 3904 bp in length
* 83982 86179: gap of unknown length
* 86180 86279: gap of unknown length
* 86280 89696: contig of 3417 bp in length
* 89697 89796: gap of unknown length
* 89797 92878: contig of 3082 bp in length
* 92879 92978: gap of unknown length
* 92979 96577: contig of 3599 bp in length

```

```

* 96578 96677: gap of unknown length
* 96678 10056: contig of 3979 bp in length
* 10057 100657: gap of unknown length
* 100658 103738: contig of 2962 bp in length
* 103739 103839: gap of unknown length
* 103840 106689: contig of 2851 bp in length
* 106690 106789: gap of unknown length
* 106790 109708: contig of 2919 bp in length
* 109709 109808: gap of unknown length
* 109809 113553: contig of 3745 bp in length
* 113554 113653: gap of unknown length
* 113654 116605: contig of 2952 bp in length
* 116606 116705: gap of unknown length
* 116706 120077: contig of 3372 bp in length
* 120078 120177: gap of unknown length
* 120178 122520: contig of 2343 bp in length
* 122521 122620: gap of unknown length
* 122621 125560: contig of 2940 bp in length
* 125561 125660: gap of unknown length
* 125661 128466: contig of 2806 bp in length
* 128467 128566: gap of unknown length
* 128567 130902: contig of 2336 bp in length
* 130903 131002: gap of unknown length
* 131003 134453: contig of 3451 bp in length
* 134454 134553: gap of unknown length
* 134554 137254: contig of 2701 bp in length
* 137255 137355: gap of unknown length
* 137356 140707: contig of 3353 bp in length
* 140708 140807: gap of unknown length
* 140808 143398: contig of 2501 bp in length
* 143399 143498: gap of unknown length
* 143499 146274: contig of 2776 bp in length
* 146275 146375: gap of unknown length
* 146376 148466: contig of 2092 bp in length.

```

FEATURES

Source

1. 148466
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-361K18"

BASE COUNT 39949 a 31869 c 32369 g 40737 t 3542 others

ORIGIN

Query Match 12.9%; Score 845.6; DB 2; Length 148466;
Best Local Similarity 75.4%; Pred. No. 3.5e-190;
Matches 1499; Conservative 0; Mismatches 360; Indels 128; Gaps 31;

```

QY 4536 GAGATTAACAGTTCTGCTGTTGAGCAGACGCTTTCTGAGAGGAGGAGCCAC 4595
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72227 GAGACGACAGAGTGTCTGCTATTGAGCAGAGGCTCTTTGTCAGATGACACCC 72286

QY 4596 CTGAAGAGAGAGCACAAGAGAGCTCCTGAGACTTGAGAGCTTGGAATCATTCTGAG 4655
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72287 CAAGAAGAGAGAGCACAAGAGAGATGCTCCAGACACTGAGGCTTGAGACCCATCTGAG 72346

QY 4656 GAGCGTGACAGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4712
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72347 GATGCTGACAGAGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 72406

QY 4713 CCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4772
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72407 CCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 72465

QY 4773 AGACTTGTGCTGTAATGACAGAGTGCAGCAATCTTGAGCTCCAGCTGCTGCTGCTG 4832
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72466 TGCG-CTGTGTTAGAAATACACAG-----CCGAGAGCTGCTGCTGCTGCTG 72506

QY 4833 CCACATGACAGATGATATTCGAATTCGAATGATCAGACAGAGTGTGATCTCTG 4892
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72507 CCAGCTTTCGGATATCCTCAAAATTCAGAGATG--TAAAGGTGATTTTCCCTG 72563

QY 4893 CTGACATTCATACGCTGAGAGATTTTAAATCTCTCTCTCTCTCTCTCTCTCTCTCT 4952
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72564 GGACATCTTCATGCTCGAAGAG--GTTCCATATCTTCTCTGAGATTCATCCAGC 72621

```

OY	4953	CCACACTGATGTCATTTGGATTAATGAAAGAAACCTGGACATTCAGTACTGGGGGCT	5012
Db	72622	CCTAACCCCTCGATTCATTTTGGAGATGAGAAAGAAAGCTGACCTTTGAAGTACAGG----	72676
OY	5013	TTAACCACCAAGGAAGACAAAGAAAGAAACATGAATTCCTTGGACTAGCTGCTGTCCA	5072
Db	72677	-----AAGACAGACAGATATGAATCTCTTAGACTA-TGTCTGTGTCCA	72719
OY	5073	CTGTGTTACATGTCCTCCCTTTTAAAAAAAATGAGTTTAAATTTTGTTCAGAGAG	5132
Db	72720	CGTGTGTTCAACATCCCT-----TTAAATATTAACCTTAAAGATGTGTCCAGAGG	72769
OY	5133	TAAAT-ATATATCCATTTTAATGATTTACAGT-----ATTATTTTAAACCTTAAGTAGGTT	5186
Db	72770	TAAATGTACATCTATTTAGTATATATATTTAATATATCATTTTGAACCTTTAAGTAGATTT	72829
OY	5187	GCCACGCTGG-CTTCTGAAAAAACCAATATATGCCGACAGGGGTGGCCACACCAAGAAAG	5245
Db	72830	GCCACGCTGTGTTCTGAAAAAACCAATATATGCCAAGAAAGGTTTGTCTGTGCCAACA--	72887
OY	5246	CGGGAAGACCTGGCTGTGACCCCTGGCTCCCATGTCTCTGTGCTCACCGGCAAGTG	5305
Db	72888	---GAAGACCTGGGCAAGTATCTGTCTCCATGTCTCTTTGGTCTGGCCAGTAAAGG	72944
OY	5306	CCCTATCTGTGAAGTATGAATATGTAGCCAAATTA-----ATACCAAGACACCTCATCTGC	5360
Db	72945	CCCTATCTCCAGAACTATCAATATGTTACCCAAATACTTATGATACCAAGACTTCTCACCTGC	73004
OY	5361	TCCTTCCCCAGCTGGATGGGGTCTCTGTAAACCTGTTTGCACATGGCCAGGGAGGCA	5420
Db	73005	TCCTTCCCCAGTATGGGTGGAGGCTCTATATGAAGC---TACACGGTGGCCAGGGAGTGA	73060
OY	5421	CTAGACACCTGTGTCTCAGCTGTGACCCCTTATGAGGACGAGACGGTGTCAATTGGCCAGTCT	5480
Db	73061	CTTGGACTTGTGTGTCTGTCTGTGACCCCTGCT- GAAGCAGGATGGGTGTTTGAACACT	73119
OY	5481	GTCCTGCTTCATTTAGATGGATGGCAAAACCCATTTTAAAG--TTATATTTCTTTGATT	5537
Db	73120	GTCCTGCTTCATTTAGATGGATGGGCGCAACCCCTGCTCTAGCTTTGTTTGTGTTTGAAGT	73179
OY	5538	TTTGTATTTTATGAGGTGTAG-----GTTTGTGTTTGTGTTTGTGTTTGTGTTT	5585
Db	73180	TTTGTATTTTATGAGTTTATGAGCTTTTGTGTTTGTGTTTATTTTGTGTTTGTGTTTGTGTTT	73239
OY	5586	TTAAGAAACATTTTATACTGGATGAGATTGCAGTGAAGACAGCTTGGATGTTGGAGC	5645
Db	73240	TATTTTAAAGAAACATTAACCTGATAGCATACAGTG-AAGCAGCTTGGGGTGTGGAGT	73288
OY	5646	TATATGCCAGCTTT-TATATCTCTCTTTCAGACAGCTCCCTTATATGAATTTGGCATTA	5704
Db	73299	GAATATCCAGCTATTTGTGATTTCTTTCAGACAGCTCCCTCTGCTGGATTTGGACACCA	73358
OY	5705	GGGATTAACAAAGCCTTTTAAAGTATTAAGATTAAGAAACCTGGTTTGGACATGGCAGCC	5764
Db	73359	GAGAGT-----ATGTCTTTAAACATATAT-AGAGGTACACAACTTGGTTATATCAGTACGCC	73413
OY	5765	TTTTCGAAGCAGGTTAGTACACCAAGAACTAACCTCCAAAGTGAGCTTTATGGACGCTGCATA	5824
Db	73414	TTTTCGAAGG-TGGTTAGTACACCAAGAGGCGATGTGAATGGCTTTGTGGATGCTGCACA	73472
OY	5825	TAGACAAGGCTTAATGTATGACACCATCTGCTCAGACAGCTGCTATTTAACCCTTAAATGACT	5884
Db	73473	TGGAAAAAACCTTAAGATGACATATGTCTGTTCATACAGCTGCGATTAGCCCTTGAATGACT	73532
OY	5885	G-AAATGACCCCTCCACTCTATTTTGTGTTT-GTTTGCACAGACTCCGGAAGAGGAG	5942
Db	73533	GAATAATGACCCCTCCGCGTCTATTTTGTGTCTTTTGTGAGGAGCTTCAGAAAAAGGAG	73592
OY	5943	GCTGCCAATCTAGTAGTACTCAATATGTGAGAACTGCTGGTCTTGGATTTTTCAT	6002
Db	73593	GCTGCCAATCTCAAGTAGTACTCAGATGTGAGGAGCCTGCTGATCTTGGG---TTTTTTCAT	73649

OY	6003	TAAATTTACGCTCATCAATTGTGCATCGTGAATAAGCATAAATGGCTTCAAAATTTTTAAAGT		6062
Dd	73650	TAAATTCAGCTCATCATATATGATGATCAATAGATTAAGTAGT-GCCTCACATTTTAAAGAT		73708
OY	6063	GGAATTCGACGTCTTTTTCACATGTAATCAAACAATGTCAAGTGCTTTATTTAAATATTCCT		6122
Dd	73709	CAAAATTCGACGTCTTTC-TCACTGTGCCAAACAATGTCAAGGCGCTTTATTTAAATATTCCTCN		73767
OY	6123	TCTGTATCATGGCATTGTGCTCTGCTGTAATACAAATTCGAAATTCAGATTTGGCATTTT		6182
Dd	73768	TCTGGGCCAGGSCATTTTGCTTACCATATGATATACACTGTCACTTACTGTTTATATATCA		73827
OY	6183	ACATGTAAATATGCAAT-----TAATTCGCAAGTTTTTAATTAATAGGCTATGGA		6228
Dd	73828	ACATGTAAACGTATTAATATGTAATGTATGCACTATTTGCGAGTTTTTATTCATATAGGCTATGGA		73887
OY	6229	CCTCATGTGCATATAGAAAAGACAGAATCTAGCTTACCAAGAAGTTGCCAACAAATGTTATTC		6288
Dd	73888	CCTCATGTGCATATATAAAAAGACAAAACCTTACTTACCACAAATTTTACAGAGTGTATTAAC		73947
OY	6289	TAGACATTAAGTAATTTAGAACATAGSACTGCTATCTCACATGGTGGCTGTGATGTCAA		6348
Dd	73948	TACACATTAAGTAATTTAGTAAGACACAGAGACTCTATCTCAGCTGCGCTGTGATGTCAA		74007
OY	6349	GTGCAGAAATGTAACAATTAACGTGTAATTTCTCATACTTTTGTATCTACTTGTACCTGTGA		6408
Dd	74008	GTGCAGAAATGTGA-AGTTAACTGCTGATTTCCCTTAATACCTTTGTATATCACTGTATCTGTGG		74066
OY	6409	TCTCTTTTGAAGAAGCATGTGGAGTGCTGTATCCCTTTTGTATTTTAATCAACAATAAT		6468
Dd	74067	TATCTTTTGAAGAAGACACTGCTGGAGGCTGT-CCTCTTGTATTTCTCATATCAACAATAAT		74125
OY	6469	GTACATA 6475		
Dd	74126	GTACATA 74132		

Search completed: February 18, 2003, 15:52:10
Job time : 17748 secs

BLANK PAGE

QY 812 SGNLAVALLRTWELPKKYEKLFODLDDLPSPRNMAKYRNVLSQNLQPPITLPLF 871
 Db 925 AANGTAVYLRKWTGQLPEKYELECEKRLSD--HLKKYQERLSTIN--PPCVPFEG 980
 QY 872 VIKKDLTLEHGNDKVDG--LVNFEKLRMAKEIRHVGMAVNMPLMFTFRKKWR 929
 Db 981 RYLTNHLLEGNDLLANTELINFSKRKRYAELI----- 1015
 QY 930 SLGSLSGSTINATVLDVAQGGHKRYRSGSEFLNAKLYE-----DAQARKYQY 980
 Db 1016 --GEIQYQYQNPYCLNLEST-----IROFFEQDLPFNGLSIKQMS---DY 1055
 QY 981 LSNLELEMEDEESTLQSLCEPATNTLPKNPDK-----KVPKSESPVAPRAGS 1030
 Db 1056 LVNESHLEIRRGCKTVKPKFRKWRPHILKSPGIKPRQNOTNSSKLSNSTSSVAAAAA 1115
 QY 1031 QQAQSLPPQQQPPRAHKITNOGLQVPAVSLYDSRKRVKVDLPFFGINSQALKLTSL 1090
 Db 1116 SSTATSI-----ATASASLIHASSIMDAP----- 1139
 QY 1091 SEEGSLERHKQAEEDTISNASSQLSPPTPOSSPRKGYTLAP-----SGTVDN 1139
 Db 1140 -----TAAANAGSGTLAGEQSPQHNPAFSVFAYIIPERNTSSNGSHQH 1186
 QY 1140 F--SDSGHSEIS-----SRSSIVSNSSPDSV-----PVSLHDER 1171
 Db 1187 TRPDQNGGEVSVAPRLPKKRGANVMANNSTLASASAMQVESPALPEHLPPQSLPDSN 1246
 QY 1172 RQHSVSYVTNLGCMGMERTMIERQYSLGSYARPSSEGRGLKATATVSSPTELSQ 1231
 Db 1247 PRASDTLAPRSP-----PKLVVSP--RHETGNKSP----HGRQNSPHTASTATVTLT- 1295
 QY 1232 DQGDRASLDAADSGRSWTSSCSSGSHDNIQTIOHQRWETLPFG-----HNHFVSGD 1284
 Db 1296 -----GMSTSGGEERCAGGYFN--SAHQQPGAVLPISPHNVPMATNMEYRAV 1342
 QY 1285 PAGL-----WASSSHMDQIMS-----DHST--KYNRQNS 1313
 Db 1343 PRPLPRRKERTESCADMAKQARDAPRLPRPDGELSPRPRLNHSSTISYLRQSHG 1402
 QY 1314 RESLEDAQSRASWASTGVGEDS--EGDGTGKRRGKQDVSTLEAS-----S 1360
 Db 1403 K-----SKEFGNSLLPLPNTSSIMIRNSAIEKRAATISOPNOAAGPIS 1448
 QY 1361 LVSVTTEET---KVPMPANLAVASSTTKGLARKEGKRYREPRPR-----PGYIGAP 1410
 Db 1449 TPLVTVSQAVATDEPLRLPISPAASSTTTS-----PLPRAMPSPMLPISHP 1496
 QY 1411 IDPFRGHANPRAKRPDDYNVALORSRMVARSPTAGRSSQQRNGHRTSSRPVKNQGMK 1470
 Db 1497 VESTSSSYAH-----QLRMQOQOQOTPRALYSQNHQNHATLPL--NRPQNH 1542
 QY 1471 PMSDEPLAP 1480
 Db 1543 SNPQGRSSP 1552
 RESULT 5
 PC1114
 SKCDD25 protein - yeast (Saccharomyces kluyveri) (fragment)
 C:Species: Saccharomyces kluyveri
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
 C:Accession: PC1114
 R:Rigouzy, T.; Gonzalez, E.; Broek, D.
 A:Title: Identification and analysis of a DNA fragment from Saccharomyces kluyveri that
 A:Reference number: PC1114; MUID:92354938; PMID:1644315
 A:Accession: PC1114
 A:Molecule type: DNA
 A:Residues: 1-1095 <PRT>
 A:Cross-references: GB:M62964; NID:g171186; PIDN:AAA34479.1; PID:g171187
 i:Gene: SKCDD25

Query Match	4.2%	Score 325.5;	DB 2;	Length 1095;
Best Local Similarity	21.6%	Pred. No. 1.1e-06;		
Matches 158;	Conservative 110;	Mismatches 240;	Indels 225;	Gaps 27;

[illegible]

A:Residues: 1-1048 <MIO>
A:Cross-references: EMBL:Z73121; NID:g1360186; PID:e245452; PID:g1360187; MIPS:YLL016w
A:Experimental source: strain S288C
R:Goffeau, A.; Purnelle, B.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64761
A:Accession: S64764
A:Molecule type: DNA
A:Residues: 1-1048 <GOF>
A:Cross-references: EMBL:Z73121; NID:g1360186; PID:e245452; PID:g1360187; MIPS:YLL016w
A:Experimental source: strain S288C
R:Purnelle, B.; Goffeau, A.
Submitted to the EMBL Data Library, April 1996
A:Description: The sequence of 32 kb on the left arm of yeast chromosome XII reveals 14 mly and a new ABC transporter homologous to the human multidrug resistance protein.
A:Reference number: S69380
A:Accession: S69390
A:Molecule type: DNA
A:Residues: 1-1048 <PUR>
A:Cross-references: EMBL:X97560; NID:g1297003; PID:e238680; PID:g1297014
R:Mogea, T.; Zimmerman, F.K.
Yeast 12, 693-708, 1996
A:Title: Sequence analysis of the CEN12 region of *Saccharomyces cerevisiae* on a 43.7 kb e conductivity regulator protein CTR.
A:Reference number: S70557; MUID:96405918; PMID:8610043
A:Accession: S70559
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1048 <MIW>
A:Cross-references: EMBL:X91488; NID:g1495203; PID:CAA62775.1; PID:g1495207
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1995
C:Genetics:
A:Gene: SGD:SCD25; SDC25
A:Cross-references: SGD:S0003939; MIPS:YLL016w
A:Map position: 12L
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology
F:744-995/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 4.28; Score 321.5; DB 2; Length 1048;
Best Local Similarity 19.98; Pred. No. 1.7e-08;
Matches 226; Conservative 166; Mismatches 402; Indels 339; Gaps 45;

QY 13 MGOEKHSLEADFTKHLTDS--LHPQVTVSSSHSGCSTSDSGSSLSLDIYQATESEA 70
175 LNQHMSMDLPRQMKAVAGSAGIYAENIDELPASKQSTSCSSSTSHSPAPQRRRGRT 234
DB 71 GMDLGLPETAVDSEDDDEEDIERASDPLMSRDIVRDLCKDPIDRTDDIEQLLEFM 130
235 IFSNVSG-----SSDEPTIMSKRKRPPLNETL--SLVRARKQDLGKLKQIMKSA 285
QY 131 HOLPAPANMTASVARELCAMVAVAVNR-AGTIVLNDGEELDSKSYILN-----G 179
286 NE--YLSNANPNSKMLNFENFTEYEVSGTIIPIIDLENLDTIPLNRLGLEDENRF 342
QY 180 SVEATYDGAELT-----CMGNSFGVSPMPDKEVKKGVKRTVVDCCOFYCINQO- 229
343 DEVDVAIDDEDEELFKHLSLSLSTLSDYFMN----KQYHADV-----VKFLIVAOHL 391
DB 230 -----DYCRILNQVEKNMQKVEEGEIIYVWKEHRLDRTGTRKHGHI 272
392 TLEDPEVFPSPMNDLPTGYEPMKPSNLINDMKDK----- 427
QY 273 KGTSEKRLTHLVEHSAVDPTFIEDPLITRTFLSPMEVKGKILKEMFNDPSLRDKVTRV 332
428 KNGSQNDIDEEDEYEPDP-----DSLILFHNLLINDSDND--LKRPFN-----LAHV 474
QY 333 VLLVNHFPNDFEGDPMATRFLEFENN--LEREKMGCHRLINIAACAARLRMLTLC 390
475 FKSKSDYFD-----VLKALIEFVQMLILER-----NLMTYA-----ARMKKNITTE 517
QY 391 PSREAPLPFTLLGSGSEKFGFIYDVSDSGSKATEAGLRGQDILVNGONFENIOLSKAM 450
DB 518 -----LLLRG-EEGYG-----SYDGGEMAEKS----- 538

QY 451 EILRNNTHTLSTVKNLFEVKEKELLTRLSEKRNAGAPHLPKITGDIKKARSYIPDLAYVE 510
539 -----DTNA-VYADSDTKNDKEMRSQVKLP-----RT----- 565
QY 511 QVIGLEKVRNKRKANTV-GGRNKLKLLDKTRISILPQKPYNDIGOSODDSVGLRQY 569
566 -----LQREYSELIMGSNRNRIKGSKHALLSYLDNENKDLFF-----DITFLITF 612
QY 570 KHIPALP-VSGTLLSSNP-----LLOSHRILDFSAT----- 602
613 RSIFTTPEFLSYLISQYNLDPPEDLCFEENYEWTKLIPVKCQVWEIMTFEQRQYWEFG 672
QY 603 PDDPDQYRFAKDDQGSRYIMISKDTAKREVYQAIREFAVTAPPDQYSLCEVSPREGV 662
673 YDEPD--LATLNDYFQVAKIKENTGVSLELEVNOKF-----KHGN 713
QY 663 IKORRLPDQLSKLADRIQLSGRYLKNMNET--ETLCSDDAOELRESOISLPLSTVE 720
714 IQENTAP-----MKTLDQIQODHSGTLYSTTE-SILAVDPVL 751
QY 721 VAFQLSRNELEFRNIEPTVEYIDLEK--LRSKTSCANLKRFEVINQETFWVASEILR 777
752 FAFQTLTLEHEIYCEITIFPOLQIKWKNTKYSYGASPLGMEFISFANKLITNFTSYVK 811
QY 778 ETNOLKRMKTIIRKIKALICRECKNNSMFAISGLNLAVPALRTTWKLPKRYEKL 837
812 EADSKRAKLSHFIFAEYCRKFNFSSMTAIIISALYSPIYLEKTLTWOAVIPOTDRL 871
QY 838 QDLQDLDPSPRNMAKRYNNVNSONLQPIIPLPFVIKDLTFLHEGNSKVDGLVNEKL 897
872 QSLNKLMDPKPKNFINYNELKSLH-SAPCVPFVEYVLSDLTFDSGNP---DIYVLEHGL 927
QY 898 RMIAKETIRHVGMAVNMDDPALMEFRKKKWRSLGSLSGSTNATVLDVAOTGCHKRRVR 957
928 KGVHDEKKYIN-----FNKRSR-----LVDILO----- 950
QY 958 RSSFLNAKKIYEDAOAKRKQYILSN-LELEMEDESQITLSLCEPRTNLPKPGCKRP 1016
951 -ETIYFKTHYDPTKQRTYIECISNLENIPIHEKYOYLSLIEP-----KP 996
QY 1017 VKSETPVAPRAGSQKQKOSLPQPOOPPPAHKINGOLQVPAVSLYSPRRKVP 1069
997 RKR-----VVRNSNSNKKSQEKSRDDOTDEGKTSTKKD-REPFKQLHKTKKKAP 1044

RESULT 7
S1417
SCD25 protein (version 1) - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein L1509; protein YLL016w
C:Species: *Saccharomyces cerevisiae*
C:Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 06-Feb-1998
C:Accession: S14177; S12942; PS0040
R:Damak, F.; Boy-Marcotte, E.; Le-Roscouet, D.; Guilbaud, R.; Jacquet, M.
Submitted to the EMBL Data Library, August 1989
A:Reference number: S14177
A:Accession: S14177
A:Molecule type: DNA
A:Residues: 1-1250 <DAMI>
A:Cross-references: EMBL:M26647
R:Damak, F.; Boy-Marcotte, E.; Le-Roscouet, D.; Guilbaud, R.; Jacquet, M.
Mol. Cell. Biol. 11, 202-212, 1991
A:Title: SDC25, a CDC25-like gene, which contains a RAS-activating domain and is a di
A:Reference number: S12942; MUID:91094833; PMID:1986220
A:Accession: S12942
A:Molecule type: DNA
A:Residues: 1-737, 'A', 738-970, 'I', 972-1250 <DAMI>
A:Cross-references: EMBL:M26647
A:Note: the authors translated the codon GAG for residue 538 as Asp, GTC for residue
R:Boy-Marcotte, E.; Damak, F.; Camonis, J.; Garreau, H.; Jacquet, M.
Gene 77, 21-30, 1989
A:Title: The C-terminal part of a gene partially homologous to CDC25 gene suppresses
A:Reference number: PS0041; MUID:89306677; PMID:2545538

A:Accession: PS0040
A:Molecule type: DNA
A:Residues: 668-679, 'NPVMTIMC', 689, 'N', 691-1250 <BOY>
A:Note: the authors translated the codon GCA for residue 747 as Asp and GAT for residue 748
C:Genetics:
A:Gene: SGD:SCD25: SPC25
A:Cross-references: SGD:S0003939; MIPS:YLL016w
A:Map position: 12L
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology
F:946-1197/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 4.0%; Score 310.5; DB 2; Length 1250;
Best local similarity 26.18; Pred. No. 8.3e-08;
Matches 105; Conservative 65; Mismatches 159; Indels 73; Gaps 12;

QY 680 QLSGRYYLKNMNET-----ETLCSDEDAQELLRESOISLLOSTVEVATQLSMRNF 731
Db EVNQKFKNGNIOEATAPMKITLDQIQDHYSGTLYSTIE-SILAVDPVLFATQITILEHE 964
QY 732 LFRNIEPTVEYIDDLFK--LRSKTSKANLKRFEVINDETFWVASEILRETNOIKRMKI 788
Db IYCEITTFDCLQIKMKYTKSYGASPLNEFISFANKLTNFISSYVKADKSKRAKLL 1024
QY 789 KHFILKIHCECKNENMFALISGLNAPYARLRTTEKLPNKYEKLPQDLQDLDPBR 848
Db SHETFLAYECKNENFSSMTDIISALYSSPIYLEKTMQAVIPQTRDLOSINKLMDPKK 1084
QY 849 NMAKYRNVNLSNQLPPIPLFPIYIKKIDLPFHGNSKVDGLVNFELRIAKEIRHVG 908
Db NFIVYRNLKSLH-SAPCVPFEGVYLSDLTFTDSGNP---DYLVHGLKGVHBEKKIYN 1140
QY 909 RMAVNMNDPALMEFRTRKKKRRKSLGSLSGSTNATVLDVAQTGCHKRRVRSRSPFNARKLY 968
Db I141 -----FKRRSR-----LVLDILQ-----EIIYFKTH 1161
QY 969 EDAOMARKVOYLSEN-LELEMEESLQTLSCBPATNTLEKPNQKPKVSESPVAPR 1027
Db I162 YDFKDKRTVIECISNLENTPIHEIKQYQSLITRP-----KPRKK-----VVPN 1205
QY 1028 AGSOKAKOSLPQPOQPPRAHKINQGLQVPAVSLYPSRKVP 1069
Db 1206 SNSNKSQEKSRDQDTDEGKTSTKKD-RFPRFQHKTKKKAP 1246

RESULT 8
S28407
guanine nucleotide-exchange activator CDC25 homolog - mouse
C:Species: Mus musculus (house mouse)
C:Date: 17-Apr-1993 #sequence.revision 17-Apr-1993 #text.change 05-Nov-1999
C:Accession: S28407; S22693; B46199; S20730
R:Gen, H.: Papageorge, A.G.; Zippel, R.; Lowy, D.R.; Zhang, K.
EMBO J. 11, 4007-4015, 1992
A:title: Isolation of multiple mouse cDNAs with coding homology to Saccharomyces cerevisiae
A:Reference number: S28407; MUID:93010996; PMID:1396590
A:Accession: S28407
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1260 <CEN>
R:Margreth, E.; Vannoni, M.; Zippel, R.; Cocciotti, P.; Brambilla, R.; Ferrari, C.; Stura
EMBO J. 11, 2151-2157, 1992
A:title: Cloning by functional complementation of a mouse cDNA encoding a homologue of C
A:Reference number: S22693; MUID:92289680; PMID:1376246
A:Accession: S22693
A:Molecule type: mRNA
A:Residues: 789-1260 <NAR>
A:Cross-references: EMBL:X59868; NID:950357; PIDN:CAA42525.1; PID:950358
R:Wiel, W.; Mosteller, R.D.; Sanyal, P.; Gonzales, E.; McIntney, D.; Dasgupta, C.; Li, P.
Proc. Natl. Acad. Sci. U.S.A. 89, 7100-7104, 1992
A:title: Identification of a mammalian gene structurally and functionally related to the
A:Reference number: A46199; MUID:92353779; PMID:1379731
A:Accession: B46199
A:Status: preliminary
A:Molecule type: nucleic acid

A:Residues: 1029-1030, 'D', 1032-1224 <WEI>
A:Experimental source: fetus
A:Note: sequence extracted from NCBI backbone (NCBIN:111101, NCBIP:111102)
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homol
F:242-428/Domain: CDC24 homology <CD24>
F:1021-1257/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 4.0%; Score 309; DB 2; Length 1260;
Best local similarity 19.8%; Pred. No. 1e-07;
Matches 219; Conservative 157; Mismatches 364; Indels 364; Gaps 44;

QY 17 EKHSLPADFTKL-HLTDSLHPQVTHVSSN-----SGCSITSDSGSS----- 57
Db ERNSLDYAKSKLEELSRIMHDEVESETNIRKNLAIERMTGCGTLLDTSGTFVROGSLM 466
QY 58 -SLSIDIQATFESAGMDLSGLPETVAVSDEDDDEEDIERASDPLMSRDIY----- 107
Db 467 QMSLSEKSKSRGLGSL-----STKKEGROCF-LPSKHLIITRSGSG 510
QY 108 -----RQLEKRPIDRTD-----DIEQL-----LEFNHOLAFANMTMSV 143
Db 511 KLHITKGVISLIDCTLLDEPENLDDEAKAGAPETHELEKIGVEPKDSLPTVIVIAST 570
QY 144 RRELCAVM--VFAVVE--RAGTIVLNDGEELDSWVILNGSVEVTPD--GKAEILCMG 196
Db 571 ROEKAAWTSIDICQVDNIRCGMLMNAFEE-----NSKVTVPQMTKSDASLYCDD 620
QY 197 NSCFVSTPMDEKRYKGVYKRTKVDCCQVCIAQDYCAILNOVEKNMCKVEBGEIYVKE 256
Db 621 VDIRESKTMS-----CKVL-QIR----- 638
QY 257 HRELDRTGTRKGIIVIKGTERLTMHLEHESVVDPFIDFLITLYTFLSPMEYCKL 316
Db 639 -----YASVERLLERL-----DLRFLSIDFLMTFLHSTRVF--TNAMVYLDKL 680
QY 317 LEWNPDSLQDKYTRVYLVVNNH-----FNDEGPAMTRFLEEENNERKMGHLR 371
Db 681 INIYRKPMSAIPARSLLELFSSSHNAKLTYGDAPKSPRASR--KESPPP----- 728
QY 372 LNIACAKAKRMRMTKTPSRREAPLPFIILGSEKGFIFVDSVSGSKATAGLGRD 431
Db 729 -LAIGTSPSPRRKLSLNTIP-----ITGKALFLA----- 758
QY 432 QILEVNGQNFENIQLSKAMEILRNTHLSI-----TVKNTLFFVEKLELTRLSEKRNA 485
Db 759 -SLGSSDSYANI-----HSPISPGKTTLDGKLCMASSLPKPTPEELDVNA 804
QY 486 PHLPKIDIKKASRYSLPDLAVDEQVIGLEKYNKSKANT---VGRNKLKLIIDKTR 541
Db 805 TIPEKPELSASRKHSSDYLKEESE-----DDONHDEDENTVEVSPKSPPTPKSFLNRT- 858
QY 542 ISILPKQPYNDIGISQODSIYGLQTKHIIPALVYSGTLSSNDLQSHRILDFSA 601
Db 859 ITPEFPFNNGILMTTCRDLVNNSTLSATAPAIATAGANEGSNKEVFR--MSLN 917
QY 602 TPDLPDQVLEVRKADQOS--RYTIMISKDTAKREVVIQAIREFAVTAPPOYSICEVSYAP 659
Db 918 TG-----FSSDQRNDIKKEVIRAAITNR--VLNVRKHWTAKTTP-----FPT 958
QY 660 EGYIKORRL-----PDQSLKADRIQLSGRYLYLKNMTEITLCSDEDAQELLRESQ 710
Db 959 DDTLKRVICFELEVNHDPDLTQ-----ERKAANAIIRTLLEETTEQHSMLLEYI 1010
QY 711 ISLQLOST-----VEVATQLSMRNFELRNI-----EPREYIDDLF 746
Db 1011 LMTEGYKTEPEFENHPDLAEOLTLDDLHVLVFSIPYEEFFGQGMKAERYERTPY---- 1066
QY 747 KLRKTSKANLKRFEVINDETFWVASEILRETNOIKRMKIHKFIALHCECKNENS 806
Db 1067 ---MKTT---KHFNVSN---FISSEIIRNEDISARASALEKWAVAVDICRCHLNHYA 1115
QY 807 MFAIISGLNAPYARLRTTEKLPNKYEKLPQDLQDLDPSSRMMAKYRVNLS-ONTOPP 865
F:1021-1257/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

```

Db 1116 VLEITSSINRSALFRLKKTWLVKSKOTSLSLDKLOKLVSSD---GRFKULRESLRNCDDP 1172
OY 866 IIPLEFVIRKDLTFLEHGNDISKV-DGLVNFELKRLMAKEIRHVQ--RMASVNMDDPALMR 922
Db 1173 CVPYLGMYLTLDFVEEGFPNTEDGLVNFMSKRMISHLIRETRQOQTYTIDP-----1227
OY 923 TRKKWRSLSLSQSGSTNATVLDVAQTGCHKRRVRRSSFLNAKKLYEDAOMARKYKQYLS 982
Db 1228 -----OPKVIOYLL 1236
OY 983 NLEEMDESLOTLSIOCEPATNT 1006
Db 1237 DESFMLEDESLYESLLEPKLP 1260

RESULT 9
525716
Ras guanine nucleotide exchange factor son-of-sevenless (sos) 1 - mouse
N:Alternate names: probable ras activator
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C:Accession: S25716; S21391
R:Bowtell, D.; Fu, P.; Simon, M.; Senior, P.
Proc. Natl. Acad. Sci. U.S.A. 89, 6511-6515, 1992
A:Title: Identification of murine homologues of the Drosophila son of sevenless gene: pc
A:Reference number: S25714; MUID:92335328; PMID:1631150
A:Accession: S25716
A:Molecule type: mRNA
A:Residues: 1-1336 <BOM>
A:Cross-References: EMBL:211574; NID:954134; PIDN:CAA77662.1; PID:954135
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; pleckstrin ref
F:459-561/Domain: pleckstrin repeat homology <PLK>
F:793-1036/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 4.0%; Score 307.5; DB 2; Length 1336;
Best local similarity 19.6%; Pred. No. 1.3e-07;
Matches 224; Conservative 169; Mismatches 409; Indels 339; Gaps 47;

OY 236 NOYENKMQVEEGEL-VWVKEHRE--LDRTGRKGIHVIKGTSERLTMHLEHESVDP 292
Db 257 NOVENFISRVIDHLSVLLGHIEDTVEMTDEGSPHPLVSGCEDLA-----ELAFDP 311
OY 293 --TFIEDFLLT--YRTFLSSPMEVGRKLL-----EMFNDPSLRDKVTRVLLVNNHEND 343
Db 312 YESYARDILRPGHGFHLSQLSKPAALYLQSIGEGFKE-AVQYVLRLLAPVNYCLHY 370
OY 344 FEGDPMATFLFEFENNLREKMGHL-RLN-----IACAkakRLMLTLTKPSREA 395
Db 371 FE--LLKOLEKSEDOEDKECKQAITALLVQSGMERIKCSLSLAKRRL-----417
OY 396 PLPFIILGSEKGFIFVDSVDSGSKATEGLRGDQILEVNGQNFENIOLSKAM-ELIR 454
Db 418 -----SESACRYSOOM-----KGQOLAIKMNEI-QKNIDGEGKDIGCCNEFTM 463
OY 455 NNTHLSTVSK--TNLFVKELLTRISEERKNGAPHLR-----KIGDI 494
Db 464 EGIITFVGAKEHNRHIFLPGLM--ICCSNHNQGRPLRGASSAEYRLKEKFPMKVQINDK 521
OY 495 KASAR--SIRPLAVDEOYI-----GLEKVKKSKANTVGGKNNKLLDKTRISILQ 547
Db 522 DDTSEKNAFEIILKDGNSVIFSAKSAEERKMMMALISLQYRSTLERMD-----572
OY 548 KPVNDIGIGOSODSIVGRQTKHIFPALPVSGTSSSPNDLLOSHNRL-DESATPDLR 606
Db 573 -----VTVIOEKEKEQMRPLSAVYRFAERD--SEENILFEENVPKAG 614
OY 607 DOVLR--VFKADQOSRYIMISKDTAKEVYQAIREFAVTATPDQOISCEVSVPEGYI 663
Db 615 IPIKAGTVKLIERLFTYHMYADP-----NFVRFLT-----YRSCRPQELLSLI 662
OY 664 KORRLPDQLSKADRIQLSG-----RYLKNMETFLGS-----D 699
Db 663 ERREIPEPEPTADRIALINGDOPLSAEILKRFREKEYIOPVOLRLVNLVCRHWENHNYDFE 722

```

```

OY 700 EDAOELLRESQ-----710
Db 723 RADDLQRRNEEFIGTVRGKAMKRWESITKIIQKRIADNGCHNITTOSSPPTVENHI 782
OY 711 -----ISLQSTVEAVATOLSMRNFELFRNIEPEYIDDLFKLRK-TSCANLKREE 761
Db 783 SRPGHIEFDLLTLHPILRIARQLTLESDDLVRAGVSELVGSWTKEDKEINSPLIKMI 842
OY 762 EVINQETFWASIILETQOLKRMKIIKHFIKALCRECKNNNSMFAIIISGLNLPVAR 821
Db 843 RHTNLTLMFEKICIVETENLEERVAVVSRIELIOWFOELNNGVLEAVSANNSSPYR 902
OY 822 LRTTWEKLPNKYEKLFODLODDLP--SRNMAKYRWLVNSQNLOPILPFPYIKKDLTF 879
Db 903 LDHTEQIIPSRQKILIEEAHELSEDIHKYTLAKRSI-----NPPCVPEFGIYLTNLIK 956
OY 880 LHEGNDISKVGLVNFELKRLMAKEIRHVGRMASVNMDDPALMFTRKKKWSLSQSGST 939
Db 957 TEEGNP-----EVLRRHGKEL-----INFSKRRRVAEITGEIQOYN 993
OY 940 NATVLDVAQTGCHKRRVRRSSFLNAKKLYEDA-----QAKRYKQYLSNLEEMDESLO 994
Db 994 QPYCLEVEP-----DIKRFENLPMGNSMEKEFTDYLFNKSLER-----1033
OY 995 TISLOCEPA-TNTLPKPNQDK--PVKSE-TSPVAPRAGSQKASLPQPOQOPPAHK 1049
Db 1034 -----EPHAKPLPRFP--KTSYPLKSEGVPRSPNRETMH---PTPLOEP---1077
OY 1050 INQGLQVPVAVSLVSRKKVPVKDLPFGINSPOALKIILSSEGLERKKQAEPTISN 1109
Db 1078 -----RKISYSRIP-----ESDETSTYASA 1096
OY 1110 ASS--QLSPPTSPQSSPRKGTTLAPSGTVNDFSGHSEISRSISYNSFSFDPVPS 1166
Db 1097 PMSPRPLTPPPASGTSMTDWCSPDS--DHSAPFHSRSASVSISLSKGTDEVPVP 1153
OY 1167 LHDERQRHSVSIVETN---LGMGRMERTMIEPDQYISLGSVPMSEGRGLVATATVIS 1223
Db 1154 PVPYPRRRESAPAESPSKIMSKHLDSPATPRPTSKAYSP---KISIDRTISID 1209
OY 1224 P 1224
Db 1210 P 1210

RESULT 10
A37488
Ras guanine nucleotide exchange factor son-of-sevenless (sos) 1 - human
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 18-Nov-1994 #text_change 18-Mar-1997
C:Accession: A37488
R:Charlton, P.; Camonis, J.H.; Gale, N.W.; van Aelst, L.; Schlessinger, J.; Wigler, M.
Science 260, 1338-1343, 1993
A:Title: Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2.
A:Reference number: A37488; MUID:93262494; PMID:8493579
A:Accession: A37488
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1333 <CHA>
A:Note: sequence extracted from NCBI backbone (NCBIP:132148)
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; pleckstrin
F:442-544/Domain: pleckstrin repeat homology <PLK>
F:776-1019/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 3.9%; Score 303.5; DB 2; Length 1333;
Best local similarity 19.0%; Pred. No. 2.1e-07;
Matches 221; Conservative 175; Mismatches 397; Indels 371; Gaps 47;

OY 236 NOYENKMQVEEGEL-VWVKEHRE--LDRTGRKGIHVIKGTSERLTMHLEHESVDP 292
Db 240 NDVENFISRVIDHLSVLLGHIEDTVEMTDEGSPHPLVSGCEDLAELADPYESYA 299

```



```
QY 287 HSNVDFPFIEDFLTYRTFLSSP-----MEVGKKLLEMFNDPSLRDKYTRVYLLVNNH 340
Db 300 RDLIRGFGFHDRL-----SQTSPGALLYLOSISEGFE-----AVQVLRRLRLAPYHC 350
QY 341 FNDGSDPAMTRFLFEFENNLEREMGHL-RLLN-----IACAKAKRRMLMTLKPS 392
Db 351 LHFEE---LLKQLEKSEDEDECKEQATLALNVOSMEKICSNLSAKRRL-----400
QY 393 REAPRLPFIILGSEKGFIVDSVDSGSKATGKRGDOILEVNGONFENITOLSKAM-E 451
Db 401 -----SESACRFYSQOM---KKQOLAIKMMEI-OKNIDGEGKIDIGCCNE 443
QY 452 ILKNHHLSTVTK-TNLFEVKELLTRLSEKNGAPHL-----KI 491
Db 444 FIMEGILTRGAKAHERHIFLFDGLM-ICCKSNHGPRILPGASNAEYRLKEFFMKVQI 501
QY 492 GDIKKASRY--SIPDLAVDEOVI-----GLEKVNKSKANTVGRNKLKIDKTRIST 544
Db 502 NDKDDTNEYKHAFEIILKDNYSYIFSAKSAEKKNNMMAALISQYRSTLERMLDVTM--- 558
QY 545 LPQKPYNDIGIGSQDSDSYGLAQTKHIIPALVYSGTSSNDLLOSHRIIDFSATPD 604
Db 559 -----LQEEKEDOMRLPSADYRREAPD--SEEMIFEEENMQPK 595
QY 605 LPDOVLK---VFRADOOSRYIMISKOTTAKAVYIOAIRFAVATAPDOYSLCEVSTPEG 661
Db 596 AGPIIKAGIVIKLEILYTHMADP-----NFVRTLTI---YRFSCKQBELSL 643
QY 662 VIKORRLPDOLSKLARIQLSG-----RY-----665
Db 644 IIEFEIPEPEPTADRIALENGQPLSAELKFRKREYIQVOLRVLVNCRHWHHHFYD 703
QY 686 -----YKKNME-----TETICSDEDAQELRES-----709
Db 704 FERDAYLLQRRKEEFTGVGRKAMKKNVESITKIIQRRKIADNGPGHNITFOSSPTVEW 763
QY 710 -----QISLQSTVEVATQLSMRNELFRNIETPEYIDLEFKLRK-TSCANLKR 759
Db 764 HISRPGFIETFDLTLHPIELARQLTLESDLYRAVQPSYELVSGVWVKEDKEINSPMLK 823
QY 760 FEEVINOETWVASELIRETNQLRKMKIKHFIKIALHCECKNFNSMFAIISGLNAPV 819
Db 824 MIRHTTTLTLMFEKCIETELERVAVUSRIEIIQVFOELNMFNVLEVASMSNSPV 883
QY 820 ARLTMTKEKLPNKKEKLFQDLOLDFP--SRNKAERYNLNSQMLQPIIPLFVYIKDL 877
Db 884 YRLDHTFEQIPSRQKTLLEAHLESDHYKKYLAKLSI---NPPCVPEFGIYLTNI 937
QY 878 TFLHEGDSKYDGLVNEFKELMLAKEIRHVGMAVNMDEPALMFRTRKKRWRSLSLSOG 937
Db 938 LKTPGEMP-----EVLKRHGKL-----INFSKRKYVAELTIGELIQY 974
QY 938 STNATVLDVAQTGCHKRRVRRSSFLNAKKLEYDA---QMARVKOYLSNLELMEDEES 992
Db 975 QNQCPLCLVES-----DIKREFENLPMGNSMEKEFTDYLFNKSLIEIPRN 1020
QY 993 LQTLSTLOCERATNTLPKNPGDK---PVKSE-TSPVAPRAGSOQKASLQPOQOPRAH 1048
Db 1021 -----PKPLPRFP--KRTSYPLKSPGVRPSNPRGTMRH---PTPLQOEP--- 1060
QY 1049 KINGLOVPAVSLYSRKKYVVKDLPFGINSQALKKILSLSEGSLEHRHKAQEDTIS 1108
Db 1061 -----RKISYSRIP-----ESTESTAS 1078
QY 1109 NASS---QLSSPTSPSSPKKGYTLAPSGTVDNFS--DSGHS-----1146
Db 1079 APNSPRLPLPFPASGASS-----TTDVCYVEDSDSHSPFSSNDTVEIQVTL 1127
QY 1147 ---EISSRSSTVSNSSPDSVYSLHDERQHSIVETN---IGMGRMERRTIEEDQY 1200
Db 1128 HGPRASVSVSISLTKGIDEVVPPVPRRRPESAPASSPSKIMSKHLDSPPAIPRPP 1187
QY 1201 SLGSYAPMSEGRGLYATATVYISSP 1224

RESULT 11
S30356
CDC25 protease homolog - yeast (Candida albicans)
C.Species: Candida albicans
C.Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 29-Oct-1999
C.Accession: S30356
R.Golberg, D.; Matbach, I.; Gross, E.; Levitzki, A.; Simchen, G.
Eur. J. Biochem. 213, 195-204, 1993
A.Title: A Candida albicans homolog of CDC25 is functional in Saccharomyces cerevisia
A.Reference number: S30356; MUID:93238685; PMID:8477693
A.Accession: S30356
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1333 <GOL>
A.Cross-references: EMBL:M94160; NID:g170834; PIDN:AAA34329.1; PID:g170835
C:superfamily: SH3 homology; CDC25-type guanine nucleotide exchange activator homolog
F:37-69/Domain: SH3 homology <SH3B>
F:1064-1305/Domain: CDC25-type guanine nucleotide exchange activator homolog <SOS>

Query Match 3.9%; Score 302; DB 2; Length 1333;
Best local Similarity 19.3%; Pred. No. 2.5e-07;
Matches 213; Conservative 157; Mismatches 338; Indels 394; Gaps 44;

QY 23 ADFTKLHLDLSLHPQYTHS-----SSHSGCITSDSGSSSLSDIYQAT 66
Db 432 ASFAQIYINGILHLSVHSLSEGFNEGRLGYDMKLNRRSSSSAPQSPASSLSTRGSS 491
QY 67 E-----SEAGMDLSGLPETAVSDEDDDEDEIE-----RASDELMR 104
Db 492 DDSTRAQKLSQDRNSGNDMNYINQLY--EIDNREVNSTIVKIFLKSANKKIKNS 549
QY 105 DIVRDLCKPDRIDTDDIEQLLEFMHQL-PAF-----ANMTSVRELCAYWF 153
Db 550 DY-----DSSADSDGEGDFRIDLPQVPRFLVDFNGMNCNPFSTKNTVLNVSGD 602
QY 154 AVERAGTIVLNGEELDSVTLNLSVETYPDGKAEILCMGNSGVSPTMCKEYMKGV 213
Db 603 DLNRHFTKIIIDHSAYDLS-----QYVDKI 629
QY 214 MTRVDDCOFVCIACODYCRILNVEKMKQVEEGEIVAKKEHRELDRTGTRKGIHVIK 273
Db 630 ---VDACE-----NILEALDPKVQNFYUENMLNENMTQD-----662
QY 274 GTSERLT-MHLVEHVSVDPTFIEDFLLYRTFLSSPMEYVKKL---LEMENDPSLRDKV 329
Db 663 ---LRLTYKSLYHCSAMVD--LIESF--DFTVFCSVKRHTGNAIDTEDESEYENPS----- 710
QY 330 TRVYLLVNNHNFDFED-PAKTRFLFEFENNLEREMGHLRLNINACAKAKRRMLMTL 388
Db 711 ---WGDDHYDNLSPDYVPLEF-----RLKQELHDLVAKITMTQSS---LTL 753
QY 389 TKPSREAPLPFIILGSEKGFIVDSVDSGSKATEAGLRGDOILEVNGONFENITOLSK 448
Db 754 EDPE-----VF-----KGLKEBEPD-----FYNEISK 776
QY 449 -----AMEILRNN-----THLSITVYTNLFEVKELLTRLS 478
Db 777 IPREKALLLSLILKEQLSEFKDGAISLNPDTLLSGYLVEIAATYTVTLITQOLI--- 832
QY 479 EKRNGAPHLPKIGDIIKAKARYISIPDLAVDEOVIQLEKVNKSKANTVGRNKLKIID 538
Db 833 EERTITINAYTRY---MODNFVQLLVERNNTSSEKADDS--YVGGHKK-----880
QY 539 KTRISILPQKPYNDIGIGSQDSDSYGLRQTKHIIPALVYSGTSSNDLLOSHRIIDFSATPD 595
Db 881 -----SDVPRWYLGDDDEYELL-----LDVKGNIKGSKRALKVSHLTHHL 920
QY 596 ILD--FSATPDLPDOVLVFRKADOOSRYIMISKOTAKAVYIOAIRFAVATAPDOYSLC 653
```

```
Db 921 SLDSNFA-----VF-----LLMFSSMMLGELLISLLARFNIE----- 954
OY 654 EYSVPEGY-----IKORLPDQSLK-----A 676
Db 955 ----PPEGLSYEEYMLWMSKKRNPIRLRYINIMKLLLEKNWSMYNEPVLRLMTFAHS 1010
OY 677 DRIOLSGR-----YYLKNMETETLCSDDED-----AQELLRESOISLLOS 717
Db 1011 DOVOYIYSLGNLNVNLERLREGERIVERDPVYIPNTKPPAPRLTKGSSLLKRPV--MDID 1068
OY 718 TWEVATQISMRNFELFRNIETPEYIDDLFKLS--KTSKANLKEEVINOETFWVASEI 775
Db 1069 YVELARQILTLREFKLYCKITFKACLAKVWGKSGLSSESIDSTQIFKASNOQLTFVGYMI 1128
OY 776 LMETNOLKMKKIIKFIKIALHCRECKNNSMFALISGLNLAPVARLRTTWELPKKYEK 835
Db 1129 LKKAPDKKRVQIIRFYIYADCKROYNFSMVAIIISALYSSEPIHKLKTWEWMADALS 1188
OY 836 LFQDLODLDPSPRNMAKYNNVLSQNLQPPILPLFPVKKDLFTLHEGNDKSV---DGLV 892
Db 1189 NLKNNMKLNNSSRNFEYRDVLKFTIGSE-PCVPFFGVYISDLTFYVHGMPDLYXNTRQY 1247
OY 893 NFEKLMIAKEIRHNGR-----MASVNDPALM 920
Db 1248 NFAKRAKTEIYSGIDRFYTGYNFOEVEPEIQKFLDAMFEKCPYIDEQYQISLNEP--- 1304
OY 921 FTRKKKMSLGSLSGSGSTNAT 942
Db 1305 -----REOPAGASNSNSTNAT 1321
```

RESULT 12

```
JC7736
C3G protein, long type - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 14-Dec-2001
C:Accession: JC7736
R.Zhai, B.; Huo, H.; Liao, K.
Biochem. Biophys. Res. Commun. 286, 61-66, 2001
A:Title: C3G, a guanine nucleotide exchange factor bound to adapter molecule c-Crk, has
A:Reference number: JC7736; MUID:21378139; PMID:11485308
A:Contents: 373-L1 adipocyte
A:Accession: JC7736
A:Molecule type: mRNA
A:Residues: 1-1086 <ZNR>
A:Cross-references: GB:AF348669
C:Comment: This protein, a guanine nucleotide exchange factor, through the interaction
```

```
Query Match 3.8%; Score 295; DB 2; Length 1086;
Best Local Similarity 23.5%; Pred. No. 4,1e-07;
Matches 150; Conservative 82; Mismatches 205; Indels 202; Gaps 26;

OY 491 IDDIKASRYSPDLAVDEQVIGL-EKYNKSKANTVGGRNKLKLLDTRISILPORK 549
Db 535 VGD-----FSYVELAGDTEKPPRLPEKKNKMLAYM-----QLLDYSPQPSMFGYQRP 583
OY 550 YNDIGIGGQ-----DDSIVGLRQTKHT--PTALP----- 577
Db 584 QSEHIYQOKNKMLMEYVGFSESGSDSSTOELAPRALPORKOLRPPSGKDGHPRPDSY 643
OY 578 --VSGTLSSSN----- 591
Db 644 SSSSGKDSKNGERSPKSLDGLSASQSEEVDELSDIHNEIMARLTLOEGDGDGVNRG 703
OY 592 SHHRIIDFSAT-PDLPDQVL-----RVFKADQ-----QSRVYIMS--KDTTAKEV 633
Db 704 GSGDILLVHATEIDRKDLVLYCEAFLLTYRTFISPELIIKKIQYREKSPFADTKKRV 763
OY 634 VIOAIEPAVTATPDQYSLCEVSVTPGEVIAIKORLPDQSLKADRIQLSGRIYLKNNMET 693
Db 764 SKNTF--FVLVRVDE--LCLVELTEE-ILK-----LIMEIIVRLVCSGSELSTARVLR- 811
```

```
OY 694 ETLCSDEDAOELLRESQI-----SLQSLTVEVATOLSMNFELEFNIEPT 739
Db 812 KNILDKVDOKKLLRCHASHQPLAARGVAARPGTLHDFHSHETLAEQQLTLDAELFYIEHP 871
OY 740 EYIDDLF-KLRKTSKANLKEEVINOETFWVASEILRETNOLKMKIIKFIKIALHC 798
Db 872 EYL--LMAKDQNEKSPNLTQFTEHFNMNSYVWRSIIMIOEKADQDERLLIKIKIMKH 929
OY 799 RECKNNSMFALISGLNLAPVARLRTTWELPKYKRLFKRODLOLDFDPSRNMAKYNNVLS 858
Db 930 RKLNNNSYLAIIALSADSAPIRLE--WQR--QTSGLAEYCTLIDSSSSFPAYRAALS 984
OY 859 SQMLQPPILPLFPVKKDLFTLHEGNDKSVYDGLVNEKLMIAKEIRHNGRMAVSMDPA 918
Db 985 --EVEPPCIPYIGLITQDLTFVHGNPDYIDGKVN----- 1018
OY 919 LMFRTKKKMSLGSLSGSGSTNATVLDVAQ--TGCHKRRVRSRSPFNAKKLYEDAQARK 976
Db 1019 -----SKRMQGE-----NILDSMRCFOGAHYETIRNDIIN----- 1049
OY 977 VKOYLSNLELMDDESLOTLSLOCBPATNTLPKNPGDK 1015
Db 1050 --FNFDFSDHLAEALWELSLIKIKRPNITRRKTDREK 1085
```

RESULT 13

```
A38985
nucleotide exchange factor CDC25 - human
N:Alternate names: Ras-specific guanine nucleotide-releasing factor
C:Species: Homo sapiens (man)
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 05-Nov-1999
C:Accession: A38985; A46199; I58371
R.Wei, W.; Broek, D.
submitted to Genbank, December 1994
A:Description: Cloning and analysis of the full length human cdc25 cDNA, a ras-specif
A:Reference number: A38985
A:Accession: A38985
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1275 <WE1>
A:Cross-references: GB:I26584; NID:9433719; PIDN:AA58417.1; PID:9433720
A:Wei, W.; Mosteller, R.D.; Sanval, P.; Gonzales, E.; McKloney, D.; Dasgupta, C.; Li,
Proc. Natl. Acad. Sci. U.S.A. 89, 7100-7104, 1992
A:Title: Identification of a mammalian gene structurally and functionally related to
A:Reference number: A46199; MUID:92357779; PMID:1379731
A:Accession: A46199
A:Molecule type: mRNA
A:Residues: 1047-1054,'A',1056-1112,'G',1114-1135,'C',1137-1275 <WE2>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIN:111098, NCBIP:111099)
R.Schneidhoffer, F.; Faure, M.; Fath, I.; Chevallier-Multon, M.C.; Apizou, F.; Dutrill
Oncogene 8, 1477-1485, 1993
A:Title: Identification of a human guanine nucleotide-releasing factor (H-GRF5) spec
A:Reference number: I58371; MUID:93275641; PMID:7684828
A:Accession: I58371
A>Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: mRNA
A:Residues: 787-1275 <RES>
A:Cross-references: GB:S62035; NID:9386046; PIDN:AA826881.1; PID:9386047
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homol
F:240-426/Domain: CDC24 homology <CD24>
F:1036-1272/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>
```

```
Query Match 3.8%; Score 290; DB 2; Length 1275;
Best Local Similarity 19.4%; Pred. No. 9,7e-07;
Matches 218; Conservative 162; Mismatches 393; Indels 348; Gaps 48;

OY 3 PLATPANHCWGQO-----EKHSLPADFTKL-HLTDSLHQYVHVSSH-----S 46
Db 386 PRYITLHDVLAHPHEHVERNSLDYAKSKLELSRIMHDEVSETEIRKNLAIERMIIE 445
OY 47 GCSTSDSGSSSL--SDIYQATSEAGDMDSLGLPETAAYSEDDDDDEDIERASDPLMSR 104
```

```

Db 446 GCETLLDTSOTFVROGSLIOVPMSSEKGIIRGLSLSEKES-----ERQCF-LFSK 497
QY 105 DIV-----RDLKDPIDRTD-----DIEQL-----LEFMQ 132
Db 498 HLIICIRGSGKHLTKNGVISLIDCTLLPEPESTEBAKSGODIDHDFKIGVEPKDS 557
QY 133 LPAFAMNTMSVRELCVAVMFVAVERAGTIVLNDGEFLDSVSLVGVETVPDCKAEI 192
Db 558 P-----FTVI-----LVASSROKAMWIDISQCD-----NI 586
QY 193 LCMG-----NSFGVPTMDKEVMKGVARTV--DDCOFVCAIOOYCRILNOYEKMM 242
Db 587 RCMGLMNNAEENSKVYVPMIKRTREGTEAEKMSDASLYC-----DVIDIRF 636
QY 243 QKVEEBCIYVAKHEHRELDRTGRKGITVIGTSERLTMHVEHSVVDPTIEDLLTY 302
Db 637 SKTMNSCKVLOIR-----YASVERLERLT-----DLRFLSIDFTLFLHST 678
QY 303 RTFLSPMEYGKLLWFENPDLADKVTYVLLMV-----NNHFNDEGDPAMT 351
Db 679 RVF-TTAVVLDKLITTKRP-----ISALPARKLRLLELFAFGQNNKL--LYGEPPKS 730
QY 352 -RLEEFEN--NLEREKMGHLRLNIAACAARRLMTLTKPSREAPLPLILGG-----404
Db 731 PRATRKFSPPPLSITKTSSPSR-----RRKLISINIP-----IITGKALD 772
QY 405 -----SEKGFIFVSVSGSKATFAGLRGQILLEVNGQNFENIOLSKAMEILLNNT 457
Db 773 LAGSLSCNSNGYTSYMSAMSPFSKAT-----LDTSK-----803
QY 458 HLTSTVNTLNFVKELLRLSEERKNGAPHLPKIGDI--KKASHYSTPDLAVDEQYIGL 515
Db 804 -----LYSSSFPNKIPDE--GDTPPEKEDPSALSKOSSSVMSREES--DIDQ-----848
QY 516 EKVKKSKANTVGGRNK--LKLTLDKTRISIDOKPYNDIGIOSODDSVIGLRQTKHI 572
Db 849 --NQSDGDGTETSPTKSPPTPKSVKKNKNSSEPLFESYNN--GVWMTSCRELNNRSALSA 904
QY 573 PTALPVSGTSSNPDLOSHTLDFSAPDLPDQYLRFVKADQOS--RYIMISKOTTA 630
Db 905 ASAALINAGNEGTPNKEKTRMSLASAG-----FPPOKNGDKFVIRRAATN 954
QY 631 KEVYIOAIRFEAVATATPD---QYSLCEVYTPREGVIKORRLPDOLSKLADRIOLSGRY 686
Db 955 R--VLNVLRHVNSKHSODEFTNDELKCKVIGLEBVHMDPELLTOERKAANI-----1005
QY 687 LKNMNETTLCSDDEDAQELRESQISLQL-----SYVEVATQISMNFE 731
Db 1006 -----IRTLQOEDPG-----DNOITTEETIOMAGVKAEPFENHSALETIAEOLLTDLHL 1054
QY 732 LFRNIEPTETI--DLEFKL-----RSKTSQANLRFEEVINOETFWVASEILRETQOLKRM 786
Db 1055 VFKKIPEFEFGQGMKLEKMERPTYIMKTTHFNDSN-----LIASLIINEDIINRVS 1110
QY 787 TIKHEIKIALHCRECKNENSPFAITISGLNLAPVARLRTTWKLPNKYKLELQDLODLFP 846
Db 1111 ALEKVVAVADICRCLHNNAVLEITSSMNRSAIFRLKTKTWKVSQOTKALDLKLOKLVSS 1170
QY 847 SRNMAKYNNVLSQLOPPIPLPFYIKKDLFLHEGDSKV--DGLVFEKLMIAKTER 905
Db 1171 EGRFNLEAL--KNCDPPCVYLGMYLTDLAFIEEGPNTEDBLVNFESKRMISHTIR 1228
QY 906 HVGRAVSNMDFALMFRTRKKKRSLSLSOGSTNATVLDVAOTGHHKRYRRSSFNLAK 965
Db 1229 EI-----RQOQOAY 1238
QY 966 KLYEDAQARKVQOYLSNLEEMDESLOTLSLOCEPATNT 1006
Db 1239 KIEHOA---KVTOYLLDOSFVMDSESLYESSLRIEPLRT 1275

```

```

guanine-nucleotide-releasing protein - rat
N:Alternate names: CDC25 protein homolog
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C:Accession: S29083
C:Shou, C.; Farnsworth, C.L.; Neel, B.G.; Feig, L.A.
Nature 358, 351-354, 1992
A:Title: Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor fo
A:Reference number: S29083; MUID:92350260; PMID:1379346
A:Accession: S29083
A:Molecule type: mRNA
A:Residues: 1-1244 <SHO>
A:Cross-References: EMBL:X67241; NID:957664; PIDN:CAA47666.1; PID:957665
A:Note: The authors translated the codon GAG for residues 135 and 137 as Glu
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homo1
F:240-426/Domain: CDC24 homology <CD24>
F:1005-1241/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 3.7%; Score 288; DB 2; Length 1244;
Best Local Similarity 18.4%; Pred. No. 1.2e-06;
Matches 201; Conservative 171; Mismatches 368; Indels 350; Gaps 42;

QY 17 EKHSLPADPTKL-HLTDLSHPQVTHYSSH-----SGCSITSDSGSSSL--SDIY 63
Db 405 ERNSLDYAKSKLELSRYVHDEVESETEIRKMLAERMITECEILLDTOTFVROGSLI 464
QY 64 QATESEAGMDLSGLPETVAVSEED-----DDEEDIERASDPLMSRDIVR 108
Db 465 QVMSSEKGIINKRRLSLSLAKGEGROCFLFKHLICTRGSGSKHLTKNGVIS---LI 521
QY 109 DC-LEKDPIDRTDDIEQLIEFM-----HOLPAFAMNTMSVRELCVAVFAVFE-- 157
Db 522 DCTLDDPEMMDDGQGEVDHDLFKIWEPEKSPPTVILVASSQOEKAAMTSDIQCV 581
QY 158 ---RAGTIVNDEEELDSVSLVNGSVETYPD---GKAELICMGSFVSPMTDEKYM 211
Db 582 DNIRCMGLMNAEE-----NSKYVPMITSQDSLICDDVDYDFSKTMS-- 627
QY 212 GVMRTKVDCCFVCIQAQDYCRILNOVEKMMQKVEEGELIVYKHEHRELDRTGRKGIT 271
Db 628 -----CKVL-QIR-----YAS 637
QY 272 IKTSERLMHLVBEHSVNDPTFIEDFLTYRTFLSPMEVGKLLWFENPDLRKYVR 331
Db 638 VERLERLT-----DLRFLSIDFLNFTLHSHRVF--DVAVVLDKLSTYKRPITAIAPS 691
QY 332 VLLMNVNH-----FNDEGDPAMTRFLEFENNLEREKMGHRLNIAACAARRLM 386
Db 692 LELLFSSHTKLLYGAPKSPASR--KFSSPP-----LAIGTSSPVARRKL 738
QY 387 TLTKPSREAPLPLILGSGSEKGFIEVDVSDGSKATEAGLRGQDOLLEVNGQNFENIOL 446
Db 739 SLNIP-----IITGG-----748
QY 447 SKAMEILR-----NTHLSI-----TVKTNLFVKELLRLRSEKRNKAPHLPKIG 492
Db 749 -KAEELASLQCPDSDGYTNHSPISPEKTTLTDSKLCVASSLIRTEPE-----IDMT 799
QY 493 DIKASRYSIDLAVDEQVYIGLEKVNKKSKANTVGGRNKLKRLDKTRISILPORYND 552
Db 800 TLRESSGFRKP-----TSDLIKESDDQSD-----VDDTEVSPTPKSPFRN 841
QY 553 IGIGOSODDSIVGLRQTKHPTALPVSGTSSSNPDLOSHTLDFS-----601
Db 842 -----RITQEFPLEFNYNGIMTGR--DLMDSNRSPLSATSAFAIATAGANE 886
QY 602 TPDLPDQVLRV-----FKMDQOS--RYIMISKOTTAKEVYIOAIRFEAVATATPD-----648
Db 887 SPANKEIYRRMSLANGTYSOQRNIDKEFYIRATNR--VLNVLRWVTKHSODEFTD 944
QY 649 --OYS-LC---EVSVTPREGVTKORRLPDOLSKLADRIOLSGRYLKNMNETTLCSDEDA 702
Db 945 LKTKVIGCFEEVYHMDLRLPQERKKAANIMRTLTQEFIEINH---SMDELMLMTEGV 1000

```


BLANK PAGE

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2003, 21:04:11 ; Search time 95 seconds
(without alignments)
3251.207 Million cell updates/sec

Title: US-09-911-826A-2

Perfect score: 7721

Sequence: 1 MKPLAIPANHGVCMQCEKHS.....PYQSQGFSTEDEDEQVSAAV 1499

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7721	100.0	1499	4	O9Y4G8
2	4626.5	59.9	1601	4	O8TEU7
3	4473.5	57.9	1509	4	O96PC1
4	4060	52.6	1391	4	O8TEU6
5	3871.5	50.1	1204	4	O9UHV4
6	3572.5	46.3	1113	4	O8TEA3
7	2358.5	30.5	834	11	O8R3E5
8	2352.5	30.5	1551	5	O9VWF3
9	2352.5	30.5	1573	5	O9SVI8
10	2173	28.1	579	4	O43386
11	1798.5	23.3	1347	5	O95WR8
12	1798.5	23.3	1470	5	O21218
13	1779	23.0	1311	5	O95NL8
14	1021.5	13.2	528	11	O8VCK0
15	677	8.8	876	11	O8VCC8
16	676	8.8	881	4	O95398

17	676	8.8	884	11	O92IC8	O92IC8	rattus norv
18	675	8.7	881	4	O8WVNO	O8WVNO	homo sapien
19	671	8.7	881	4	O95634	O95634	homo sapien
20	666.5	8.6	993	11	O92IP0	O92IP0	mus musc
21	666.5	8.6	1011	11	O9E026	O9E026	mus musc
22	665.5	8.6	1011	4	O8WZ2A	O8WZ2A	homo sapien
23	664.5	8.6	1011	4	O95636	O95636	homo sapien
24	659.5	8.5	1011	4	O8TAA4	O8TAA4	homo sapien
25	658.5	8.5	696	11	O8VIP9	O8VIP9	mus musc
26	654.5	8.5	699	11	O9CW52	O9CW52	mus musc
27	649	8.4	861	5	O95U79	O95U79	dirosophi
28	637	8.3	957	5	O9V9A3	O9V9A3	dirosophi
29	455	5.9	96	4	O9UFT5	O9UFT5	homo sapien
30	447.5	5.8	580	4	O92565	O92565	homo sapien
31	433	5.6	436	11	O92IC7	O92IC7	rattus norv
32	431	5.6	456	4	O9UHV5	O9UHV5	homo sapien
33	392	5.1	287	11	O8RIR1	O8RIR1	mus musc
34	365	4.7	1721	5	O8SSQ0	O8SSQ0	dictyostel
35	354.5	4.6	1596	5	O8SSU1	O8SSU1	dictyostel
36	354.5	4.6	1596	5	O969D6	O969D6	dirosophi
37	327	4.2	1575	5	O8SSN3	O8SSN3	dictyostel
38	321.5	4.2	929	5	O8SSQ3	O8SSQ3	dictyostel
39	321.5	4.2	1048	3	O12037	O12037	saccharomyc
40	299.5	3.9	426	5	O8T9J6	O8T9J6	dirosophi
41	299.5	3.9	612	5	O9VNR5	O9VNR5	dirosophi
42	298.5	3.9	1359	5	O8T6G6	O8T6G6	dictyostel
43	295	3.8	1086	11	O91Z22	O91Z22	mus musc
44	294	3.8	1077	4	O13905	O13905	homo sapien
45	292	3.8	1413	5	O9NBD3	O9NBD3	caenorhabd

ALIGNMENTS

RESULT 1	
O9Y4G8	PRELIMINARY; PRT; 1499 AA.
AC	O9Y4G8; 01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	KIAA0313 protein.
GN	KIAA0313.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI	TaxID=9606;
NCBI	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=BRN;
RX	MEDLINE-97349984; PubMed-9205841;
RA	Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT	"Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";
RT	DNA Res. 4:141-150(1997).
RL	EMBL; AB02311; BAA20772.1; -
DR	InterPro; IPR000595; CNMP_binding.
DR	InterPro; IPR001478; PDZ.
DR	InterPro; IPR000651; RASGEFN.
DR	InterPro; IPR001895; RASGEFN_CDC25.
DR	InterPro; IPR000159; RA_domain.
DR	Pfam; PF00027; CNMP_binding; 1.
DR	Pfam; PF00595; PDZ; 1.
DR	Pfam; PF00788; RA; 1.
DR	Pfam; PF00617; RASGEFN; 1.
DR	Pfam; PF00618; RASGEFN; 1.
DR	SMART; SM00100; CNMP; 1.
DR	SMART; SM00228; PDZ; 1.
DR	SMART; SM00314; RA; 1.
DR	SMART; SM00147; RASGEFN; 1.
DR	SMART; SM00229; RASGEFN; 1.

DR PROSITE: PSS0042; CNMP_BINDING_3; 1.
DR PROSITE: PSS0106; PDZ; 1.
SQ SEQUENCE 1499 AA; 167416 MW; 1909B8A12637E001 CRC64;
Query Match 100.0%; Score 7721; DB 4; Length 1499;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKPLAIPANHGVMGQEKHSIPADFTKLHLTDSLHPQVTHVSSHSQCSITSDSGSSSUS 60
DB 1 MKPLAIPANHGVMGQEKHSIPADFTKLHLTDSLHPQVTHVSSHSQCSITSDSGSSSUS 60
QY 61 DIYQATESEAGMDLSGLPETAVIDSDEDDDEEDIERASDPLMSRDIVROCLEKDPIDRTD 120
DB 61 DIYQATESEAGMDLSGLPETAVIDSDEDDDEEDIERASDPLMSRDIVROCLEKDPIDRTD 120
QY 121 DDIEQLLEFPHQOLPAFANMMSVRELCAMVFAVERACTIVLNDGEELDSVILNGS 180
DB 121 DDIEQLLEFPHQOLPAFANMMSVRELCAMVFAVERACTIVLNDGEELDSVILNGS 180
QY 181 VEVYTPDGKAEILCMGNSFGVSPPTMDKEYMKGVWRTKVDDCOFVCIAQODYCRILNQVEK 240
DB 181 VEVYTPDGKAEILCMGNSFGVSPPTMDKEYMKGVWRTKVDDCOFVCIAQODYCRILNQVEK 240
QY 241 NMQKVEEGEIVWKKEHREIDRTGTRKGHIVIGTSERLTMHLVEHSVVDPTFIEDFLL 300
DB 241 NMQKVEEGEIVWKKEHREIDRTGTRKGHIVIGTSERLTMHLVEHSVVDPTFIEDFLL 300
QY 301 TYRTFLSPMEVGKLLLEWENDSLARDKVT RVLLAWNNHNFDEEGDPAMTRLEEFENN 360
DB 301 TYRTFLSPMEVGKLLLEWENDSLARDKVT RVLLAWNNHNFDEEGDPAMTRLEEFENN 360
QY 361 LERKMGHGLRLNLIACAAKARRMLTLTKPSREAPLPTLLGSEKSGFGTIVDSVSGS 420
DB 361 LERKMGHGLRLNLIACAAKARRMLTLTKPSREAPLPTLLGSEKSGFGTIVDSVSGS 420
QY 421 KATFAGLKRGDQILEVNGQNFENITOLSKAMEILRNNTHTLSTVTNTLFEVKELTRLSEE 480
DB 421 KATFAGLKRGDQILEVNGQNFENITOLSKAMEILRNNTHTLSTVTNTLFEVKELTRLSEE 480
QY 481 KRNGAPHLPKIGDIKASRYSIDPLAVDVQVIGLEKVNKSKRANTVGGRNKLKILDKT 540
DB 481 KRNGAPHLPKIGDIKASRYSIDPLAVDVQVIGLEKVNKSKRANTVGGRNKLKILDKT 540
QY 541 RISTLPQKPYNDIGIGSQSDSYVGLRQTKHPIPALPVSGTLSSNDPLQSHRRLIDFS 600
DB 541 RISTLPQKPYNDIGIGSQSDSYVGLRQTKHPIPALPVSGTLSSNDPLQSHRRLIDFS 600
QY 601 ATPPLPOVLVFKADQOSRYIMISKDTAKEVYIOAIRFAVATPDPQSLCEVSYTPE 660
DB 601 ATPPLPOVLVFKADQOSRYIMISKDTAKEVYIOAIRFAVATPDPQSLCEVSYTPE 660
QY 661 GVIKORRLPDQSLKADRIQISGRYVLLKNNMETETLCSDEDAQELRESQISLQSTVE 720
DB 661 GVIKORRLPDQSLKADRIQISGRYVLLKNNMETETLCSDEDAQELRESQISLQSTVE 720
QY 721 VATQLSMRNFELFNIEPTETIDFLKRSKTSKANLKRFEVINQETFWVASILIRETN 780
DB 721 VATQLSMRNFELFNIEPTETIDFLKRSKTSKANLKRFEVINQETFWVASILIRETN 780
QY 781 QLKRMKIKIHFIKIALCHRECKNFNSMFAISGLNLPAVALRTTWELPKYKELFQDL 840
DB 781 QLKRMKIKIHFIKIALCHRECKNFNSMFAISGLNLPAVALRTTWELPKYKELFQDL 840
QY 841 QDLFDPSSNMKYNVNLSONLQPPITPLFPVYIKKDLTFLHEGDSKVDGLVNEKELMT 900
DB 841 QDLFDPSSNMKYNVNLSONLQPPITPLFPVYIKKDLTFLHEGDSKVDGLVNEKELMT 900
QY 901 AKELRHVGRMAYVMNDPLAMTRTKKWRKRSIGLSQSGSTNTAVLDVQGTGKKRVRRSS 960
DB 901 AKELRHVGRMAYVMNDPLAMTRTKKWRKRSIGLSQSGSTNTAVLDVQGTGKKRVRRSS 960
QY 961 FLNAKKLYEDAQMARKVQYLSNLELEWDEESLOTLSLOCEPATNTLPKNPDGDKPVKSE 1020

DB 961 FLNAKKLYEDAQMARKVQYLSNLELEWDEESLOTLSLOCEPATNTLPKNPDGDKPVKSE 1020
QY 1021 TSPVAPRAGSQOKAOSLPQPOQPPPAHKINQGLVPAVSLYPSKKKVPKDLPPFGINS 1080
DB 1021 TSPVAPRAGSQOKAOSLPQPOQPPPAHKINQGLVPAVSLYPSKKKVPKDLPPFGINS 1080
QY 1081 PQALKITLSEEGSLERKKAQEDTISNASSQLSPPTSPQSSPRKGYTLAPSGTVNFE 1140
DB 1081 PQALKITLSEEGSLERKKAQEDTISNASSQLSPPTSPQSSPRKGYTLAPSGTVNFE 1140
QY 1141 SDGSHSEISSRSIYNSNFFSDVPVSLHDERQRHSVSIVETNLGMRMERTMTIEPDQY 1200
DB 1141 SDGSHSEISSRSIYNSNFFSDVPVSLHDERQRHSVSIVETNLGMRMERTMTIEPDQY 1200
QY 1201 SLGSYAPMSEGRGLYATAVISSPSTEELSDQGDRAVSLDAADSGKSWTSCSSGSHDI 1260
DB 1201 SLGSYAPMSEGRGLYATAVISSPSTEELSDQGDRAVSLDAADSGKSWTSCSSGSHDI 1260
QY 1261 QTIQHRSWETLPEGHTHFDYSGDPAGLWASSHMDQIMFSDHSTRKYNQONQRESLEQA 1320
DB 1261 QTIQHRSWETLPEGHTHFDYSGDPAGLWASSHMDQIMFSDHSTRKYNQONQRESLEQA 1320
QY 1321 QSRASWASTGYWGEDSEGDTGTRKRGKDVSEIASSSLTSVTTEETKVPMPAHIAV 1380
DB 1321 QSRASWASTGYWGEDSEGDTGTRKRGKDVSEIASSSLTSVTTEETKVPMPAHIAV 1380
QY 1381 ASSTYGLARKKEGRYREPPPPGYITGPTDFPBGSHHPARKPPDYVALQSRMVAR 1440
DB 1381 ASSTYGLARKKEGRYREPPPPGYITGPTDFPBGSHHPARKPPDYVALQSRMVAR 1440
QY 1441 SSDTAGSSVQOPHGHPTSSRPVKNPOMHKPNESDRLAPYOSQSGSTEDEDEQYSAV 1499
DB 1441 SSDTAGSSVQOPHGHPTSSRPVKNPOMHKPNESDRLAPYOSQSGSTEDEDEQYSAV 1499
RESULT 2
Q8TEU7 PRELIMINARY: PRT: 1601 AA.
ID Q8TEU7
AC Q8TEU7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Rapi1 guanine nucleotide-exchange factor PDZ-GEF2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuiperij H.B., de Rooij J., Reihmann H., van Triest M.,
RA Wittinghofer A., Bos J.L., Zwartkruis F.J.T.;
RT "Characterization of the PDZ-GEFs, a family of guanine nucleotide
exchange factors specific for Rapi1 and Rapi2.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF478468; AAL79915.1;
SQ SEQUENCE 1601 AA; 179406 MW; E69D208431909A0E CRC64;
Query Match 59.9%; Score 4626.5; DB 4; Length 1601;
Best Local Similarity 64.1%; Pred. No. 1,4e-258;
Matches 959; Conservative 163; Mismatches 299; Indels 75; Gaps 26;
QY 20 SLPADFTKLHLTDSLHPQVTHVSSHSQCSITSDSGSSLSDIYQATESEAGMDLSGLP 79
DB 165 SLPADFTKLHLTDSLHPQVTHVSSHSQCSITSDSGSSLSDIYQATESEAGMDLSGLP 79
QY 80 ETAVASDEDDDE-EDIERASDPLMSRDIVROCLEKDPIDRTDDIEOLLEFPHQOLPAFAN 138
DB 225 ETAVASDEDDDEEDIEDIR-TDPLQGRDLVRECLEKEPADKTDIDIEOLLEFPHQOLPAFAN 283
QY 139 MTMSVRELCAMVFAVERAGTIVLNDGEELDSWVILNGSEVYVTPDGKAEILCMGNS 198
DB 284 MTMSVRELCAMVFAVERAGTIVLNDGEELDSWVILNGSEVYVTPDGKAEILCMGNS 198


```
QY 199 FGVSPTMDKRYMGVMTKVDCCOFCVICAODYCRILINOVKKNMOKVEEGEIVMVEHR 258
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 344 FGITPFLDKOYMGIVRTKYVDDCOFCVICAODYWRILNHNVEKNTHKVEEGEIVMVEHR 403
QY 259 ELDRGTGRKHGIYIKGTSERLTMLHVEHNSVDPTEIEDFLTYRTFLSSPMVEGKLE 318
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 404 ELDRSGTGRKHGIYIKATPERLIMHLEHNSIVDPPTYIEDFLTYRTFLSSPMVEGKLE 463
QY 319 WENDPSLRDKVITVILVYNNHNFDEGDPAMTRFLDEEPNNIERKMGHLELTIACA 378
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 464 WFTIDSLRDKVITVILVYNNHNFDEGDPAMTRFLDEEPNNIERKMGHLELTIACA 523
QY 379 AKKRRLMTFTKPSREAPLPIFILLGSEKGFJFVDSVSGSKATGAPKRGQIIEVNG 438
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 524 AKKRRLMTFTKPSREAPLPIFILLGSEKGFJFVDSVSGSKATGAPKRGQIIEVNG 583
QY 439 QNFENTOLSKAMEILRNHNLITVKTNLVFEKELLTRLSEKRNAPHLPIGDIKKAS 498
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 584 QNFENTPMKAVEILRNHNLITVKTNLVFEKELPR--TEQKSGVPHIPKIAK-KKS 641
QY 499 RYSPDLADVDEOYIGLEKYNNKSKANTV-GGRNKLKILDKTRISILPOKPIYNDIGIG 557
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 642 RHSIOHVPDIEQF--SQEKGSKRYKANTAGGRNKIRKIDKTRFSLPDKFLSDGLSQ 700
QY 558 SODDSIVGLRQTHIPALPVSGTSLSSNPDLQSHHRLIDFSATPDLPPQVLRVFKADQ 617
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 701 SODDSIVGRHCHSLAIMIPGTLSSSDLDLPPTSMIDFSNPDPDIPQVLRVFKADQ 760
QY 618 QSRVIMSKDTAKENVVQAIREFAVATPDPQYSLCEVYTPRGVIAKORRLPOLSKLAD 677
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 761 QSCYIISKDTAKENVVHAEHGLGASDPTSLCEVYTPRGVIAKORRLPOLSKLAD 820
QY 678 RIQSGRYIKNNMETETLCSDEDAQELRESQISLQSTVEYATOLSKRNLPRNTE 737
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 821 RIQNGRYIKNNMETETLCSDEDAQELRESQISLQSTVEYATOLSKRNLPRNTE 880
QY 738 PTEYIDDLFLRSKTCANCFREEVNOCETFWVASIELRETQOLKMKILIKFIKIALH 797
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 881 PTEYIDDLFLRSKTCANCFREEDIVNOETFWVASIELRETQOLKMKILIKFIKIALH 940
QY 798 CRCKENNSFAIISGLNLAPVARLRTTWKELPNKYEKLQDLODFPSPRNMAKYRNL 857
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 941 CRCKENNSFAIISGLNLAPVARLRTTWKELPNKYEKLQDLODFPSPRNMAKYRNL 1000
QY 858 NSQNLQPIIPLPVYIKDLTFLEHGDSYVDGLVNEFKLMTAKELRHVGRNASVMDP 917
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1001 NSQNLQPIIPLPVYIKDLTFLEHGDSYVDGLVNEFKLMTAKELRHVGRNASVMDP 1060
QY 918 ALMRTRKKKWRSLGSLSGOSTNATVLDVAOTGGHKKRVRRSSFLNAKLYEDOMARKY 977
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1061 ALMRTRKKKWRSLGSLSGOSTNATVLDVAOTGGHKKRVRRSSFLNAKLYEDOMARKY 1111
QY 978 KOYLSNLELEMEDESLQTLQCEPATNTLPKNFGDKRPYK-SETSPVAPRAGSOQKAS 1036
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1112 KOYLSNLELEMEDESLQTLQCEPATNTLPKNFGDKRPYK-SETSPVAPRAGSOQKAS 1171
QY 1037 LPOPOQPPRAHKITNGLQYPAVSLYPSRKKVPYKDLPPGINSPOAKLITLSIEGSL 1096
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1172 LPOPOQPPRAHKITNGLQYPAVSLYPSRKKVPYKDLPPGINSPOAKLITLSIEGSL 1219
QY 1097 ERHKAQEDTSSMSQSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1156
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1220 ERHKAQEDTSSMSQSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1277
QY 1157 NSSPDSVPSLHDERROHSHSVIETNLGMGRMERTMIEPDQYSLSYAPMSBG----- 1211
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1278 NSSPDSVPSLHDERROHSHSVIETNLGMGRMERTMIEPDQYSLSYAPMSBG----- 1330
QY 1212 RGLYATATVISPSTEELSDQDGRASILDADSGRSGWTSKSSGSHDNIOIIOHRSWET 1271
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1331 RGLYATATVISPSTEELSDQDGRASILDADSGRSGWTSKSSGSHDNIOIIOHRSWET 1387
```

```
QY 1272 L-PFGHTHPDYGDPAGLWASSSHMDQIMFSDHSTKYNRONOSRESLEQOASRASWAST 1330
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1388 LNSYRHTHD---DPI---AEVPTDESEYSCSKSRFCGCKSLG---RKSMTSSS 1437
QY 1331 G1WGEDSESDGTGIRKRGKDVSIKESLSTVTTTEETKPPVMPAHIVASSTTKGLIA 1390
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1438 ST-SDYEPNNGYVKKR-----VLESTPAESSEGLDKRDAIDPYKTVSTSTEGGLIV 1489
QY 1391 -----RKEGRYREPPPTPGYIGIPITDPREG-HSHPAKRPDPYVVALORSRMYRSSD 1443
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1490 YCVTSRKDDRYREPPPTPGYIGISLADKEGPHN--LKPDPYVAVORSKMHNSUS 1547
QY 1444 TAGPSSVQOPHGHPSTSPVYKPKQWHPKPNESDRLAPYOSQGFSTEDDEDEQYSAV 1499
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1548 RLPPASLSSNLVACVSKYITQOPQRHNLQPFHKKLDGVDA--DSRDEDEQYSAV 1601

RESULT 3
Q96PCL PRELIMINARY: PRT: 1509 AA.
AC 096PCL:
DT 01-DEC-2001 (TREMblrel, 19, Created)
DT 01-DEC-2001 (TREMblrel, 19, Last sequence update)
DT 01-JUN-2002 (TREMblrel, 21, Last annotation update)
DE Rap guanine nucleotide exchange factor.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-11524421;
RA Gao X., Satoh T., Liao Y., Song C., Hu C.D., Kariya K., Kataoka T.;
RT Identification and Characterization of RA-GEF-2, a Rap Guanine
RT Nucleotide Exchange Factor That Serves as a Downstream Target of M-
RT Ras."
RL J. Biol. Chem. 276:42219-42225(2001).
DR EMBL: AF94782; AAK8368.1; -.
DR InterPro: IPR000595; CNMP_binding.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGEFN_CDC25.
DR InterPro: IPR000159; RA_domain.
DR Pfam: PF000027; CNMP_binding; 1.
DR Pfam: PF000595; PDZ; 1.
DR Pfam: PF00788; RA; 1.
DR Pfam: PF00617; RasGEF; 1.
DR Pfam: PF00618; RasGEFN; 1.
DR PROSITE: PS50042; CNMP_BINDING_3; 1.
DR PROSITE: PS50106; PDZ; 1.
SQ SEQUENCE 1509 AA: 169527 MW: 726CDBA71B359976 CMC64;

Query Match 57.9%; Score 4473.5; DB 4: Length 1509;
Best Local Similarity 66.2%; Pred. No. 8.9e-250;
Matches 919; Conservative 150; Mismatches 257; Indels 63; Gaps 23;

QY 20 SLRADFTKLTLDSLHPOYTHVSSSHSGCSITSDSGSSLSIDYQATRESAGMDLSGLP 79
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 165 SLRADFTKLTLDSLHPOYTHVSSSHSGCSIASDSGSSLSIDYQATRESAGMDLRLP 224
QY 80 ETAVDSEDDDE-EDIRASDPLMSRDIVRDCLEKDPIDRTPDDIDBOLLFFMHQLPAFAN 138
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 225 EGPVDSDEDEDEBEIDR-IDPLOGRLVYECLEKEPADXTDDIDBOLLFFMHQLPAFAN 283
QY 139 MTMSVRELCAVNVFAVYERAGTIVLNDGEELDSWVILNGSVLEVYTPDKAKELICMNS 198
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 284 MTMSVRELCSVMFEVVEQAGAILLEDQELDSWVYILNGTIVISHPDQKVENLFPNGNS 343
QY 199 FGVSPTMDKRYMGVMTKVDCCOFCVICAODYCRILINOVKKNMOKVEEGEIVMVEHR 258
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 344 FGITPFLDKOYMGIVRTKYVDDCOFCVICAODYWRILNHNVEKNTHKVEEGEIVMVEHR 403
QY 259 ELDRGTGRKHGIYIKGTSERLTMLHVEHNSVDPTEIEDFLTYRTFLSSPMVEGKLE 318
```

```
|||||
404 ELDSSGRKGIHIVKATPERLIMLIEHSHVDPTIEDFLTTRTLESLDVGIKLLE 463
319 WFNDSLRDKATRVVLLWNNHNFDEGDPAMTRFLEEFENNEREKMGHLLRLNTACA 378
464 WFKIDSRLDKVTRIVLLWNNHNFDEGDPAMTRFLEEFENNEREKMGHLLRLNTACA 523
379 AKAKRRLMTLTKPRREAPLPIPLILGSGSEKGFIVDSGSKATENGKRGDQILEVNG 438
524 AKAKRQVAVLOKASRESPLQSLNGSEKGFIVEGVEPSKADGSLKQDQIMEVNG 583
439 QNFENIOLSKAMEILLRNTHLSITVKNLFEVKELTRLSEKRNKGA.PHLPKIDIKKAS 498
584 QNFENITFMKAVEILLRNTHLALVTKNINIFVKELLR-TEQEKSGVPHIPKIAE-KKSN 641
499 RYSLPDLAVNEQYIGLEKVNKKSKANTV-GGRNKKLILDKTRTISILPOKPYNDIGIQ 557
642 RHSHQHPGDIQOT-SQEKSGKKVKNANTVSGGRNKKIRKIDKTRFSLPRLKFSGGLSQ 700
558 SODDSIVGLRQTKHHPALPVGSLSSNPDLQSHHRIIDFSATPD-----LPDQVLR 611
701 SODDSIVGTRHCHSLAIMPGLTSSSPDLQPTSMDFS-NPSAVGFYIYIPDQVIR 759
612 VFKAQOSRYIMSKDTTAEVNIQAIREFAVTATPDQYSLCEVYVPEGVIKORLLPDQ 671
760 VFKVDOOSCYIILSKDTTAEVNFHAEVFEGLTASDYSICEVSYVPEGVIKORLLPDQ 819
672 LSKRADHIOISGRYILKNMNETELCSDEDAQELRSQSILQSLTSEVATQISNNEFE 731
820 FSKRADHIOISGRYILKNMNETELCSDEDAQELKESQISMLQSLTSEVATQISNNEFD 879
732 LFRNIEPTXYIDDLFKLRSTSCANLRFEEVINQETFFWVASEILFRNTQNLKRAKIKHF 791
880 LFRNIEPTXYIDDLFKLNSTGNTGHLKRFEDIYVQETFFWVASEILFRNTQNLKRAKIKHF 939
792 IKTALHRECKNENSMFAIISGLNLABVARLRTTWEKLPNKYEKLFODLQDLEDPSSNMA 851
940 IKTALHRECKNENSMFAIISGLNLABVARLRTTWEKLPNKYEKLFODLQDLEDPSSNMA 999
852 KYRNLNSQLOPPIIPLPVVIKKDLFEHGNDSKVDGIVNFEKLMIKETIRHNGRMA 911
1000 KYRNLNSQLOPPIIPLPVVIKKDLFEHGNDSKVDGIVNFEKLMIKETIRHNGRMA 1059
912 SVNMDPALMFRTRKKKRWISLSQSGSTNATVLDVAGTGGHKKRVRSFSLNKKLTYEDA 971
1060 SAMMDPALMFRTRKKKRWISLSQSGSTNATVLDVAGTGGHKKRVRSFSLNKKLTYEDA 1118
972 QMARVYQYISNLELDEESLQTLISQCPATNTLPKNRGDKPKYK-SETSVAPRAGS 1030
1119 QMARVYQYISNLELDEESLQTLISQCPATNTLPKNRGDKPKYK-SETSVAPRAGS 1178
1031 QOKAQSLPPOQOPPPAHKINQGLQVAVASLPSRKVKVYKDLPPRGINSPOALKLISL 1090
1179 QOKAQSLPPOQOPPPAHKINQGLQVAVASLPSRKVKVYKDLPPRGINSPOALKLISL 1227
1091 SEEGSLERHKKQADITISNASSQSLSPPTSPQSPKRYTLABSGTVNDFSQSHSEISS 1150
1228 TEEDIS--GKKHTEEDISVASSLSHSPSPASPOGSPKRGYTLIPSAKSDNLDSDSHSEISS 1284
1151 RSSTIVNSPDSVPSLHDERQRHSVIVETNLGMRGMRRTMIPPDQYSLGSYAPMSE 1210
1285 RSSTIVNSPDSVPSLHDERQRHSVIVETNLGMRGMRRTMIPPDQYSLGSYAPMSE 1337
1211 G-----RGYATATVLISSPTELSODQGRASLDAADSGRSGMTSCSSSHDNITIOH 1265
1338 GWTLKPSLILKCLAVSSSVNEBISQ---EHITTEAADSGRSGMTSCSSSHDNITIOH 1394
1266 QNSWEVL-PEGHTHEDYSGDPAGLWASSSHMQDIME-SDSHSTKYNRONQRESLEQAQOSRA 1324
1395 PKSMDFLANSYRHTHLD---DPI---AEVPTDESEYSCSKSCSRTGCGQCSLE---RK 1444
1325 SNASSGTYGGEDESGDTGTRIKRGGDVSTLEASSSLTGVTTETKPYVMPAHIAVASST 1384
|||||
```

```
Dh 1445 SWNSSSL-SDTEPNNGTYKRR-----VLESTPAESSEGLDPKADATDPVYKVTYST 1496
QY 1385 TKGLIARKE 1393
Dh 1497 EKLIEHQ 1505

RESULT 4
Q8TEU6 PRELIMINARY; PRT: 1391 AA.
AC Q8TEU6;
DT 01-JUN-2002 (TREMUREL 21, Created)
DT 01-JUN-2002 (TREMUREL 21, Last sequence update)
DT 01-JUN-2002 (TREMUREL 21, Last annotation update)
DE Rap1 guanine nucleotide-exchange factor PDZ-GEF2B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuiperij H.B., de Rooij J., Rehmann H., van Triest M.,
RA Wiltgenhofer A., Bos J.L., Zwartkruis F.J.T.;
RT "Characterization of the PDZ-GEFs, a family of guanine nucleotide
RT exchange factors specific for Rap1 and Rap2."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
EMBL: AF478469; AAL79916.1;
SQ SEQUENCE 1391 AA: 157495 MW: CC1882FA932EB618 CRC64;

Query Match 52.6%; Score 4060; DB 4; Length 1391;
Best Local Similarity 62.6%; Pred. No. 6.3e-226;
Matches 845; Conservative 143; Mismatches 215; Indels 146; Gaps 24;

QY 20 SLPADFTKHLNLSLHPDYTHVSSSHSGSITSQSSSLSDIYQATSEBAGMDLSGLP 79
165 SLPADFTKHLNLSLHPDYTHVSSSHSGSITSQSSSLSDIYQATSEBAGMDLSGLP 224
80 ETAVDEDDDE-EDIERASDPLMSRDYRDCLEKPRIDTDITQQLLEFHHQLPAFAN 138
225 EGVYDEDDDEEDIEDR-TDPLQGRDLVRECELEKPAKTDITQQLLEFHHQLPAFAN 283
139 MTMSVRELCANVAVFVERAGTIVLNDGEEDLSVILNGSEVYVPGKAEILCMGNS 198
284 MTMSVRELCANVAVFVEVQAGAILLEDQELDSWVILNGYELSHDPGKYNELFMGNS 343
199 FEGSPMDKEYMKGAVRTKVDCCQFYCIAQOQYCRILNOYENKMKOVVEBGETVMVKEHR 258
344 FGITPLDLQYMHGIVRTKYVDDQFYCIAQOQYCRILNOYENKMKOVVEBGETVMVKEHR 403
259 EDRGTGRGHVYIKGTSERLTMHLYEESHVYVDPFIEPFLTYRFLSSPMVEGKILLE 318
404 ELDSTGRGHVYIKGTSERLTMHLYEESHVYVDPFIEPFLTYRFLSSPMVEGKILLE 463
319 WFNDSLRDKATRVVLLWNNHNFDEGDPAMTRFLEEFENNEREKMGHLLRLNTACA 378
464 WFKIDSRLDKVTRIVLLWNNHNFDEGDPAMTRFLEEFENNEREKMGHLLRLNTACA 523
379 AKAKRRLMTLTKPRREAPLPIPLILGSGSEKGFIVDSGSKATENGKRGDQILEVNG 438
524 AKAKRQVAVLOKASRESPLQSLNGSEKGFIVEGVEPSKADGSLKQDQIMEVNG 583
439 QNFENIOLSKAMEILLRNTHLSITVKNLFEVKELTRLSEKRNKGA.PHLPKIDIKKAS 498
584 QNFENITFMKAVEILLRNTHLALVTKNINIFVKELLR-TEQEKSGVPHIPKIAE-KKSN 641
499 RYSLPDLAVNEQYIGLEKVNKKSKANTV-GGRNKKLILDKTRTISILPOKPYNDIGIQ 557
642 RHSHQHPGDIQOT-SQEKSGKKVKNANTVSGGRNKKIRKIDKTRFSLPRLKFSGGLSQ 700
558 SODDSIVGLRQTKHHPALPVGSLSSNPDLQSHHRIIDFSATPD-LPDQVLRVFKADQ 617
701 SODDSIVGTRHCHSLAIMPGLTSSSPDLQPTSMDFS-NPSAVGFYIYIPDQVIR 759
|||||
```

```
QY 618 OSRYIMISKDTTAKAYVIOAIRFAVATPDPQYSLCEVSVTEPGVTKORRLDPQSLKLD 677
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 761 QSCYIISIDTAKAYVFAVHEFGLTASDYSICEVSVTEPGVTKORRLDPQSLKLD 820
QY 678 RIQSGRYILKNNMETETVLCSDDEDAOELLRESOISLQSLSTEVAATQLSMRNFELEFRNIE 737
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 821 RIQLNRYILKNNMETETVLCSDDEDAOELLRESOISLQSLSTEVAATQLSMRNFELEFRNIE 880
QY 738 PREYIDDLFLKSKTSCANKLRFEEVINOETFWASSELLRETNOLKRMKILIKHFIKIALH 797
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 881 PREYIDDLFLKSKTSCANKLRFEEVINOETFWASSELLRETNOLKRMKILIKHFIKIALH 940
QY 798 CRECKNFNSMFAIISGLNLAIPYARLTTEWEKLPNKYEKLEFODLQDLEFDSRNMKAYRNV 857
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 941 CRECKNFNSMFAIISGLNLAIPYARLTTEWEKLPNKYEKLEFODLQDLEFDSRNMKAYRNV 1000
QY 858 NSCONLOPPIPLFPYVYKCOLTLFHEGNSDKVDGLVNEFKLRIMAKIRHVGNMAYNMOP 917
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1001 SSOSMOPPIPLFPYVYKCOLTLFHEGNSDKVDGLVNEFKLRIMAKIRHVGNMAYNMOP 1060
QY 918 ALMERFRKKKMWISLGSOGSTNATYLDVAQOTGCHKKRVRSSEFLAKKLYEDAOAKRV 977
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1061 AMERFRKKKMWISLGSOGSTNATYLDVAQOTGCHKKRVRSSEFLAKKLYEDAOAKRV 1111
QY 978 KOYLSMLEMBEESLOTSLQCEPATNTLPKNPGDKRPVK-SETSPVAPRAGSQCAQS 1036
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1112 KOYLSMLEMBEESLOTSLQCEPATNTLPKNPGDKRPVK-SETSPVAPRAGSQCAQS 1171
QY 1037 LPOPOQOPPAHKKINGLOVPAYSLYPSRKKYVPKDLPEFGINSPOALKILSLSEESGL 1096
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1172 LPOPOQOPPAHKKINGLOVPAYSLYPSRKKYVPKDLPEFGINSPOALKILSLSEESGL 1219
QY 1097 ERHKKQAEPTISNASSQSLSPPSSPRKGTTLAPSTGVNFSGSHSEISSRSIYS 1156
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1220 --GKTEPTIYASSLHSSPPASPOGSPHK-----VSSI-----ISDHSKKS 1261
QY 1157 NSSFDSVPVSLHDERQRHSVSIVETNLGMRMERTMIEPDQYSLGSAYPAMESEGLYA 1216
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1262 GQSCP-----GIG-----GALV---QKILQI 1280
QY 1217 T-ATVYSSPTELESDQDGRASLDAADSGRGVSCSSGSHDNIOITQHORSWETLPG 1275
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1281 TRSTAKRTSTETEKATEENDR-----TSCENTT-----RKRMSPF- 1316
QY 1276 HHMFVDSGPAGLMASSHMDQIMFS-----DHSTYVNRONOSRESLEQAQASMA 1327
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1317 -----RRLERMLSRERLWNSQKEDTDHNOATESCEKVKDVGSNIKDE 1359
QY 1328 SSTGWGEDESGDPTGT-----KRR 1347
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1360 KGSATFNSNSQGSNTLNCFTYFRKSKRR 1388
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
O9UHVA PRELIMINARY: PRT: 1204 AA.
AC O9UHVA:
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Dt 01-MAY-2000 (TREMBLrel. 13, Created)
Dt 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
Dt 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE PDZ domain-containing guanine nucleotide exchange factor I.
OC Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RA KAWASAKI H., Chen E.J., Springett G.M., Graybiel A.M., Housman D.E.;
RT "A new family of Rap guanine nucleotide exchange factors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF11947; AAF22004.1; -.
DR HSSP: Q12959; 1PDR.
DR InterPro: IPR000595; cNMP_binding.
```

```
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGEFN_CDC25.
DR InterPro: IPR000159; RA_domain.
DR Pfam: PF00027; cNMP_binding; 1.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00788; RA; 1.
DR Pfam: PF00617; RasGEFN; 1.
DR Pfam: PF00618; RasGEFN; 1.
DR SMART: SM00100; cNMP; 1.
DR SMART: SM00228; PDZ; 1.
DR SMART: SM00314; RA; 1.
DR SMART: SM00147; RasGEFN; 1.
DR SMART: SM00229; RasGEFN; 1.
DR PROSITE: PS50042; cNMP_BINDING_3; 1.
DR PROSITE: PS50106; PDZ; 1.
SQ
SEQUENCE 1204 AA: 137135 MW: 7AEDFE5CD00646B CRC64:

Query Match 50.1%; Score 3871.5; DB 4; Length 1204;
Best Local Similarity 75.5%; Pred. No. 4e-215;
Matches 756; Conservative 105; Mismatches 125; Indels 15; Gaps 8;

QY 20 SLPADFTKLHLDLSHPQYTHVSSSHSGSITSDSGSSLSLDIYQATESEAGDMISGLP 79
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 215 SLPADFTKLHLENPHQYTHVSSSGCSIASDSGSSLSLDIYQATESEAGDVLTRLP 274
QY 80 ERAVDEDDDE-EDIERASDPLMSRDYRDCLEKPIRDTDDIQLEFHHQLPAN 138
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 275 EGVVDEDEDEDEEDR-TDPLQGRDLVRECELEKPEADKTDIDDIQLEFHHQLPAN 333
QY 139 MPMSVRELCVAVVFAVYRAGTYVNDCELDPSVILNGSVYVYPPGKATILCMGS 198
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 334 MMSVRELCVAVVFAVYRAGTYVNDCELDPSVILNGSVYVYPPGKATILCMGS 393
QY 199 FGVSPTMDKEYMGVARTKYDDQFYCIAQODYCRILNOYKEMQKVEEGELVWKER 258
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 394 FGITPLDQYMHGIVRTKYDDQFYCIAQODYCRILNOYKEMQKVEEGELVWKER 453
QY 259 ELDRTRTKRGHIVIKTSERLIMHLVEHSVDPRTIEDFLTYRFLSPMAYGKILE 318
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 454 ELDRSGTRRGHIVIKATPEELIMHLVEHSVDPRTIEDFLTYRFLSPMAYGKILE 513
QY 319 WFNPDRLRQVTVLVLMVNNHFNDEGDPAMTRFLPEEFNNLERKMGHRLNLNAC 378
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 514 WFNPDRLRQVTVLVLMVNNHFNDEGDPAMTRFLPEEFNNLERKMGHRLNLNAC 573
QY 379 AKAKRLMLTLKPSRAPLPFTLLGSEKGFGLFVDVSDSGSKATAGLRKQDILEVNG 438
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 574 AKAKRQVYLQKASRESPLQFSLNGSEKGFGLFVGVERGSKAASGLRKQDILEVNG 633
QY 439 QNFENIQLSKAMETLRNNHLSITVTNLFVFKELLTRLSEKRNKAPHLPTGDIKKAS 498
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 634 QNFENITFMKAVELIRNNHLSITVTNLFVFKELLTRLSEKRNKAPHLPTGDIKKAS 691
QY 499 RYSPIDLVADVEQVIGLEVYKSKRANTV-GGRNKKLTKLIDKRISTLPOKPVNDIGIO 557
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 692 RSTQHPGDIEQT-SQEGSKKVKANTVSGGRNKKLTKLIDKRISTLPOKPVNDIGIO 750
QY 558 SQDSDIVGLRQTKHIPALPVSGLTSSSNPDLLQSHHRLIDFAGTADLPDQVLRVAKDO 617
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 751 SQDSDIVGRHCHSLAIMPITGLSSSSPDLLQPTSMDFSNPDIDPQVLRVRYKVD 810
QY 618 OSRYIMISKDTTAKAYVIOAIRFAVATPDPQYSLCEVSVTEPGVTKORRLDPQSLKLD 677
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 811 QSCYIISIDTAKAYVFAVHEFGLTASDYSICEVSVTEPGVTKORRLDPQSLKLD 820
QY 678 RIQSGRYILKNNMETETVLCSDDEDAOELLRESOISLQSLSTEVAATQLSMRNFELEFRNIE 737
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 871 RIQLNRYILKNNMETETVLCSDDEDAOELLRESOISLQSLSTEVAATQLSMRNFELEFRNIE 930
QY 738 PREYIDDLFLKSKTSCANKLRFEEVINOETFWASSELLRETNOLKRMKILIKHFIKIALH 797
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 931 PREYIDDLFLKSKTSCANKLRFEEVINOETFWASSELLRETNOLKRMKILIKHFIKIALH 990
```

Qy	798	CRECKNEMSAIIISGTLNAPVARIPTWKEPLPNKEKEFLQDLODLPDSRMMAVYRVL	857
Db	991	CRECKNEMSAIIISGTLNAPVARIPTWKEPLPNKEKEFLQDLODLPDSRMMAVYRVL	1050
Qy	858	NSONLOPITPLPFVIAKDDTLFHEGNDKSVQDVLNVEFKRLTAKEIRHVGMAVNDP	917
Db	1051	SSQSGOPITPLPFVYVAKDDTLFHEGNDKSVQDVLNVEFKRLTAKEIRHVGMAVNDP	1110
Qy	918	ALMRTTRKKKWRSLGSLSSQSTNATYVLDVAQTEGHHKRVRRSSFTNKKLYEDAOAMRV	977
Db	1111	AMEROR-----SLSSQSTNSMNDV--OGGAHKKRRRRSSLLNNAKLYEDAOAMRV	1161
Qy	978	KQYLSNLELEWDEESLQTLQLOEPAINTLTPKKPGGKRVK	1018
Db	1162	KQYLSNLELEWDEESLQTLQLOEPAINTLTPKKPGGKRVK	1202

RESULT 6	
Q8TEA3	
ID	Q8TEA3
AC	PRELIMINARY;
DT	01-JUN-2002 (TIREMBLrel. 21, Created)
DY	01-JUN-2002 (TIREMBLrel. 21, Last sequence update)
DR	01-JUN-2002 (TIREMBLrel. 21, Last annotation update)
DE	CDNA FLJ33738 fis, clone HEP15081, highly similar to PDZ domain-containing guanine nucleotide exchange factor 1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.
RA	Okitaani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT	"NEDO human cDNA sequencing project."
RL	Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR	EMBL; AK074318; BAB85049.1; -
SQ	SEQUENCE 1113 AA; 126109 MW; B12B7DFDE08C502B CRC64;

Query Match	46.3%	Score 3572.5	DB: 4	Length 1113
Best Local Similarity	70.1%	Pred. No. 6.6e-198		
Matches 718	Conservative 104	Mismatches 159	Indels 43	Gaps 14

QY	141	MSVARELCAMVAVFVYRACIIVLNDSEELDSMSVILNGSVLEYVPGKKLEILCMGMSFG	200
Db	1	MSVARELCAMVAVFVYRACIIVLNDSEELDSMSVILNGSVLEYVPGKKLEILCMGMSFG	60
QY	201	VSPFMDEYMKGVNRKTVDDCCQFVCIQAODYCRILNDOVERNMOKVVEEGEIVAWKEHREL	260
Db	61	ITPLIDQOYMHGIARTVYDDCCQFVCIQAODYWRILNHVEKNTKHVEEGEIVAWHEHREL	120
QY	261	DRGTGRGHVIVIGTSERLMLHLVEHSVYDPPFIEFLLITRYPLSSPMEVCGKLLWF	320
Db	121	DRGTGRGHVIVIVATPERLLMLHLVEHSIVDPPIEDFELITRYPLESPDVGKLELWF	180
QY	321	NDESLDKVTRYVVLWVWNHNFDEEGBPAMTRFEEPENLREKMGHILRLINCAAK	380
Db	181	KIDSLDKVTRYVVLWVWNHNFDEEGBPAMTRFLEEKNEEDTKMGHILRLINCAAK	238
QY	381	AKRRKLV---LITPSEAPLPFLLGGSEKSGEILYDVSVDGSKATAGLRKDDILEVN	437
Db	239	-ORLMEETGCAAGFPFVPSQFSLNGSEKSGEILYFEGVSPGSKADSGLRKDDIMEVN	296
QY	438	GGPEFNIQTSKAMEILLRNHNLSTYKTNLFYRKELLIRLSEKRNCAAPLPIGIDIKA	497
Db	297	GGPEFNIETPKAAELLRNHNLHLYKTNLFYRKELLIRFTEDEKGGVPIPIPIAE-KKS	354
QY	498	SRSISIDPLADVQVIGLEKVNKSKSANTY-GGRNKLKILDTTRISILPQRYNDIGG	556
Db	355	NHRISIGHVDPEDIGA-SDEKSGKKVANYVSGGRNKLIRKILDTTRISILPQKLFSGGJS	413

QY	557	QOSDDSVLQRTKTHPLPAVSGTSLSSNDLLOSHNRILDPGATPDLDDOYLRYPKAD	616
Db	414	QOSDDSVTRICRCSRLAIMPRLPGTLSSSSDPLDPTTSMIDRSPSDIPDOYIVRFKVD	473
QY	617	QOSRYIMISKDTTAAEVYIQALREFAPATPDDOYSLCEVSTPESGVTKQRRLDQSLKA	676
Db	474	QOSCVIIISKDPTTAAKEVFNHAEFGILTGASDGYSLCEVSTPESGVTKQRRLDQSKLA	533
QY	677	DBTOLSGRYLKNMNETETLCSDEDAQGLLESQISLLQSLTYEVATQLSMRNPELRNT	736
Db	534	DRQLNGRYLKNMNETETLCSDEDAQGLVESQISLMQSLSTIEVATQLSMRPEDFLRNI	593
QY	737	EPTETXIDDLFKLRSTSCANIKREFEYVNOETFWVASEILRETNOIKRMKIIKHFIKAL	796
Db	594	EPTETXIDDLFKLNSTGTMTLKRFPEDLYNOETFWVASEILTEANQIKRMKIIKHFIKAL	653
QY	797	HRECKENFSNFAITISGLNLAPARLRTTWELPKNYEKLEFQDLQDLFDSRMARYRV	856
Db	654	HRECKENFSNFAITISGLNLASVARLRTTWELPKNYEKLEHLODLQDLFDSRMARYRNI	713
QY	857	LNSQNLQPIIPPLPFIYKIDDLPLFHEGSDSVQDLYNFEKTLMAKIEIRYGVMAVYND	916
Db	714	LSSQSMRPPIIPPLPFIYKIDDLPLFHEGSDSVQDLYNFEKTLMAKIEIRYGVMAVYND	773
QY	917	PALMFRTKKKKWRISIGSLSGSTNATYLDVAQOTGHHKRRVRSFYNNAKLYEDAQARK	976
Db	774	PAMFRTKKKKWRISIGSLSGSTNSMIDY-QGGAHKKRRARRSLLNAKLYEDAQARK	832
QY	977	VQOYLSNLELMDDESQDLSLQCEPATNLPLKRPQGGKRYK-SETPVAPRAGSOQKKA	1033
Db	833	VQOYLSNLELMDDESQDLSLQCEPATNLPLKRPQGGKRYK-SETPVAPRAGSOQKKA	892
QY	1036	SLPQPOQPPPAHKNINGLQYPAVSLSPSRKKYVKKLPPFGINSPOALKIILISEEGS	1099
Db	893	HLHGG-----HHVQYQLPAPVANHPIRKKQGTND-PALNLSLQ---KVLGTEES	941
QY	1096	LEHHKQAEITISNASSQSLSSPPTSPQSSPRKKGYTLAPSGTYDNFSDSGHSEISRSKSI	1159
Db	942	---GKRHIEDTISVASSSLHSSPPASPOGSPHK-----VGSII-----ISDHSKI	982
QY	1156	SNS	1159
Db	983	SGS	986

RESULT 7

Q8R3E5

AD O8R3E5 PRELIMINARY; PRT: 834 AA.

AC O8R3E5.

DT 01-JUN-2002 (Tremblrel, 21, Created)

DT 01-JUN-2002 (Tremblrel, 21, last sequence update)

DT 01-JUN-2002 (Tremblrel, 21, last annotation update)

DE Similar to rap guanine nucleotide exchange factor (Fragment).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.

FR EMBL: BC025553; AAH25553.1; -.

FT NON_PBR 1 1

QO SEQUENCE 834 AA: 92545 MW: E699BBE227FE6385 CRC64;

Query Match	30.5%	Score 2358, 5	DB 11	Length 834
Best Local Similarity	59.0%	Pred. 6.8e-128		
Matches 520	Conservative 94	Mismatches 203	Indels 63	Gaps 19
QY	629	TAKEVIVIAIEEAVNTATPDQYSLCEVSYTEPGVITKORRLPDQSLKADRIOLSGRYYLK	688	
Db	5	TAKEVIVCAVDFEGLTGSDSYSLCEVSYTEPGVITKORRLPDQSFKLADRIOLNERYYLK	64	


```

115 61 Y-ASSTLEAEIDISGLVESVDS-DEEDLAEISMDSTLVRAVPCLDEKDAERSEED
123 IEOLLEHMLPFAANTMVRRELCAVWFAVERAGTIVLNDGEEDSMVSLNVSVE
116 VEVLLEBTQKATATNTTLVLRALCSVWFAVVDKAGTAVSDMSVLYLNKAVE
183 VTPDGAELICMGNSFGVSPITMDKEYKGMRTVDCOVCIACQDYCRILNQYKMN
176 IEHANGSREELQMGDSFGLPTMOKLYHRGVMRTKDDCQFVCTQTDYRIHQGEENT
243 QKVEE-GEIYWKKEHELDRTGT-----RKCHIVYIGTSERLTMHL
236 RRMDENGFVWVWTELKSGAGTDSAGSGSATGASALNMKGHVAVIGTPERLLOOL
284 VEEHSVDPFIEDELLTYRTELSPMEVGKILLEMEN-----DPSLRDKVTRVLL
296 VEEHSMRDPYVEDELLTHRFIQNPQEVTSKLLHMFDELOVDANKTOELDRVTRVLL
336 WVNHHFDFEGDPAMTRFEFEENLREKMGCHLRLLNACAKAKARRLMTLTKPSREA
356 WVNHHFDFEADYEMEFLEFEELERKKLLSOLRLHITACAKAKAMRSCTILRSSDE
396 PLPEILLGSE-----KGFIFVDSVSGSKATEAGLKRGOILEVNGQNFENIQ
416 PLNRIVGVEYELRGVAITAGNAAGIYISHVEPESKAQDVGLKRGQDIEHVNQSLDHT
446 LSKAMELRRNTHLSTVKTLPYFKELLTRLSE-----KRGAPHLR
476 SKRLELITGLTHLSIVSKNLGFKELMOALEHGCTAGSGSISAGSGSEKSVRSFRI
490 KIGDIRK-----AS-----RYSIPDLAVD-----
536 CANDIATLHGKSDTDELSVSASNRAMHVRLLSSVMDLLDQPCARPQRPVSGSGNMA
509 ---VEOYIGLEKVN-----KSKAN-----TVGG-----RMLKATILDKTRISILPQK
596 SNFMQOL--LOSVNSSAKKSGNSDDQDTKGFMTLAPKRRLQKALAKMNL-----
549 PYNDIGIGOSODDSIVGL-----ROTKHIPALPVSG-----TLSSNPDL
648 -LKNQNHGSSSLNDSSTLLNDPKSKLSAVSSCSSTOSSINGCTIVSGGGIYQOSNPDL
590 -----LQSH-HR-----ILDFSAT-----PDLPDQYLAVEKA
707 TSLVYDGSADGANGGGRGLQVNYLNAHHRPSASTLTNTNSTOHLBPDYDHVLKYKA
616 DQOSRYIMISKOTTAKEYVIOAREFAVTATPDQYSLCEVSVPTEGYIKQRRPLDQSKL
767 DOTKVVLYIKETTANHEVVMLTLOEFGIHPDSSNFSICEVSVGGGAVKORRLPDQJNL
676 ADRQLSGRAYLKKNMETETLCSDEDAOELLRESQISLLOSTVEVNTQUSMRPFELRN
827 AERISFARAYLKLNDSTETPLVPELDELAVRESNVHRLNAYELAIQTLLODFANFRQ
736 IEPTFYIDLQFLKLSKTSKANLKRFEVINOQFPVASEILRETNOUKRMKILHFTKIA
887 IESREYDELFELRSRYGVRPLSKFAELVNMEMVWVSEICAHINIVRRMKIVAQFKIA
796 LHCECKNFSMFAITISGLANLAPVRLRTTWKLPNKYEKLEFDLQDLFDPSSRMAYRN
947 RHCECKNFSMFAIVSGLGAVSRLOQTEWKLPSSKYORLFDLQDIMPSSRMMSKYRQ
856 VLNSQNL-OPRIIPFPVYKDLFELHEGNSKVDGVNPEKLLMAIKRTHVGRMAYN
1007 LVSAELLAQHAIIPFPYVVKDLFFIHLGNDTRVGDGLNFEKLLMAKEVRLTLHMSSSP
915 MD-----PALMERTRKKWRSLGSLSGOSTNA-----TVLDVAQTGCH-----KKVRSSSEFLNA
1067 YDLISLELKQSPSNLFLSNQMSASQSNAACTV--IANNAGATIKRKKKSTAAPNP
965 KKLVEDQMARKVQYLSNLELMEDESLOTLSQCEPA-----TNLPLK-----

```

```

1125 KMEEAQWVRVAKAYLNSLKLISDEDLHKFSLCECPAHGSGTYSISHGNTSHRSGG
1010 -----NFGDKPKVSETS-----PVAPRAGSQOK
1185 GSISGAGSGSSGGSSSLNAGDOLSTYSHTSSSASRANSLSLKRRPSSPTLSTSS
1034 AQSLEPPQOQPPPAHKINOGLQVPAVSLYPSRKKVPYKDLPPRGINSPOALKILSLSEE
1245 TSSTSDHQRR-----QMINNG-----PKRGTASPOAVKMKMLSLSES
1094 GSLERHKQAEEDTISNASSQLSS--PPTSPOSSPRKGYTLAPGTVDNKSDSGHSISSR
1281 SKLRPH-----QFPVPRHSGTMAGVIRPLHHMHAAHGFSTPSREGV-----TSPA
1152 SSIVSNSFSVSPVSLHDERQRHVSIVETNLGMRMERRTIIEPD-----QYSLGSYA
1327 TSAVANVQCTPRP-SFCSHRRLASGCGNIIPSRAIHERSHSDTRAPPPLPSVDLSLES--
1207 PMSEGRQATATVAVISSPSTIELSODQGRASLDAADSGSWTSCSSGSHNIOIQHQ
1384 -----SSVTT--FRDLPLRKSVTS-----GSISCDSG-----YVHQ
1267 RSMETLRFGHTHPDYSDDPAGLMASSSHMDQIMFSDHSTLKYNNQNSRESLEQAQRASW
1414 QO-----HYHLOY-----OQOQOQNSOHEPSPVY
1327 ASSTGWGEDESDDTGTRKRGKQDYSIEAESSLSVTTEETKVPMP-----AHIAVAVS
1440 TAA-----DCLRLQOISNNATVTRNL-NSPCOSTNPTPTPPRPNQPTATQLSA
1383 STTKGLIARKEGRYRE-----PPPTPGYIGIPIITDPEG--HSHPARKPPDYNV
1489 PPTAAAYMHARSQHQQLOQQOQSLAMPPEPPPY-NVP-----PLGSIYSH-----
1431 ALQSRMVARSDDTAGSSVQOPHGH
1534 -----HGGTAGSRHLNHHMHG
RESULT 9
095V18
ID 095V18 PRELIMINARY; PRT: 1573 AA.
AC 095V18;
DT 01-DEC-2001 (TReMBurel, 19, Created)
DT 01-DEC-2001 (TReMBurel, 19, Last sequence update)
DT 01-JUN-2002 (TReMBurel, 21, Last annotation update)
DE Guanine nucleotide exchange factor.
GN DIZZY OR CG9491.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Botallico A.G., He O.;
RT "dizzy, a novel Drosophila guanine nucleotide exchange factor.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL AF34686; AML28130.1; -.
DR FlyBase: FBgn0031798; dizzy.
DR InterPro: IPR000595; cNMP_binding.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR000651; RaseGFN.
DR InterPro: IPR001895; RaseGRF_CDC25.
DR InterPro: IPR000159; RA_domain.
DR Pfam: PF00027; cNMP_binding; 1.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00788; RA; 1.
DR Pfam: PF00617; RaseGFN; 1.
DR Pfam: PF00618; RaseGRF; 1.
DR PROSITE: PSS0042; cNMP_BINDING_3; 1.
DR PROSITE: PSS0106; PDZ; 1.
DR SEQUENCE 1573 AA; 171796 MW; 44881IE7621BBFIQ CRC64;

```

Query Match 30.5%; Score 2352.5; DB 5; Length 1573;
Best Local Similarity 36.2%; Pred. No. 4,1e-127;
Matches 621; Conservative 225; Mismatches 458; Indels 413; Gaps 53;

3 PLAI PANHGVGQOEKHSPLADFTKLHLDLHPQVTHVSSHSGSITSDDGSSLSLDI 62
13 PTIAGTGVVGGSSSTINRP-----ELHQKCNR--GSHSPDTSAYSGSDTMASN 60
63 YQATEEAGMDLSGLPETAVIDSEDDDEDIRASDPLMSRIYVDCLEKPIDRTDD 122
61 Y-ASSLEAEIDISGLVESVDS---DEEDLAESMDSLTVRAVDRCLEKDAEESDED 115
123 IEOLLEPMHOLPAFANMTMSRRELCAVMFAVERAGTIVLDNGEELDSMTYLNGSVE 182
116 VEVLLEFTQGLKFTNITLAVRALCSVMFAVYADKAGTYVMSDEGLDSMTYLNGAVE 175
183 VYTPDKAEILCMGNSFGVSPITMDKEYMGVMRTKVDDCQFYVCIADQYCRLLNOYEKNM 242
176 IEHANSRRELQMGDSFGILPTMDKLYHNGVMRTKDCDQFVCITQTDYRYHQGEENT 235
243 QXVEEE-GEIVMYKHEHREDTRGT-----RKGHIVYKGTSEERLTMH 283
236 RREHDEGEVYVWTELRSLIGVGTDSAGSGSATGASASLNMKRGHVJLRGTEPERLQOL 295
284 VEEHSVVDPTFIEDFLITYRTFLSPMEVGGKLEWFN-----DPSLRDXYTRVYL 335
296 VEENSMTPYVEDFLTHIFIONQEVYTKLLHMFDELOYAHKTOELRDVTRYVLL 355
336 WYNNHFNDEGDPAMTRFLEEFENNLEREKMGHLRLNIAACAARLRLMTLTKPSREA 395
356 WYNNHFTDFADYEMMEFLEVFALLERKLLQLRLHIAACAARMSRCTLTRSSRDE 415
396 PLRFEILLGSE-----KGFIFVDVSDGSKATEGLRGOIILEVNOQNEINTQ 445
416 PLNFRTVGVGELRGVAIANGNAAVGIYISHVEGSAQDVGKRGQIHEVNOQSDIHYT 475
446 LSKAMEILRNTHLSTVKTNLVFKEKELLRLSEE-----KRNAGPLP 489
476 SKRALELTCTHLSISVKSNNLGFKEIMQALHGGGTAGSGSISAGSGFVKYRSRFR 535
490 KIGDIK-----AS-----RYSIPDLAVD----- 508
536 CANDIAKLHGRSDTDELSSVASNRAHMYRLSSVMDLDPDCAPOQPVYSGSONMA 595
509 ---VEQYIGLEKN---KSKAN-----TYGG-----RKLKILDKTISTILPOK 548
596 SNFMQOL--LQSVNNSAKKSGNSNSDDQDTKGGEFTLAPKRRLQKALAKMVL----- 647
549 PYNDIGIGGODSDIVGL-----RQTKHLPALPVSG---TLSSNPDL 589
648 -LKKOHGSSSLNDSDTLLNDPKSKLSAVSSCSSTOSSINGCTVSGGGLYIOSOSNPDL 706
590 -----LQSH-HR-----ILDESAT-----PDLPOVLERYFKA 615
707 TSLNYDGGSDAGNGGGRLOVNYLNAIHBRPASAFTLTNSTQSHLLPDPDHVLYKYKA 766
616 DOOSRYIMISKDTAKENVYQAIRERAVYATFPQOYSICEVSVPEVYIKORLPDOLSKL 675
767 DOCKYVLLYKETTAAHEVAMLTLOEFGIHPSSNFSICEVSVGDGAVKORRLPDOLNL 826
676 ADRILOISGRVYLLKNNMETELFCSDEDAOELLRESOISLQLSIVEVATOLISMFELFRN 735
827 AERISPAARYLLANDSTELVDELALBELVRSNVHETLNLAYELAIQTLTDDPGFTFRQ 866
736 IEPTEYIDDLFKLSKTSYSCANLKRFEVINQETFWVASEILRETNOLKRAKIIKHFIXIA 795
887 IESTEVDELFEELRSRYGVMLSKPAELVREMFWVYSELCAHNLYRNRKIYKOFIXIA 946
796 LHRBECKNFSMAIISGLNALAVARLRTWEKLPNKYEKLFODLODLEDPSSNMAYRN 855
947 RHCKEGRNFSMAIIVSIGLGHAVSRRLQRTWEKLPKRYKQRLFNDLODLMPSRNMXYRQ 1006

856 VLNSQNL-QPPILEPVPVKKDITFLHEGNSKVDGLVNFEEKLRMTAKETIRHVGMASYN 914
1007 LVSAELLACHPIIPFPYIKKDLTFHLGNDTPVDLDLINEFKRLMAKAYRLLTHMCSSP 1066
915 MD---PALMFTRRKKRWSLGSOGSTNA---TVLDVAQTGH---KKVARRSSFLLA 964
1067 YDLSTLELKGQSPNALFESLNMOSASQSNAAAGTV--IAANAGQATIKRRKKSTAAPNP 1124
965 KKIYEDAOAKARKYKOLYLSNLELEMEDESLOTLSLOCEPA-----TNLPR----- 1009
1125 KMFEEAOVVRKAVYLNLSIKLISLDEDLHKTSLCEEPAGSGTYSGSISHVNTSHRSGG 1184
1010 -----NPDCKPVKSTSPAPRAGSOOKASOLPQOQOPRPAH 1048
1185 GSISGAGSGSGGGGGSSSLNAGDOLSTSHTS-----SSAPNSSLKRRHP----- 1234
1049 KINQGLQVPAVSLYPS-----RKKVPYKDLRPPGINSPOALKILLISEGSLERHKK 1101
1235 -----SSPFLSTSTSTSTSDHQROMHNGPKFGTASPOAVKMLLSSESKIRPH-- 1286
1102 QAEIDTSMNSQJLS--PPTSPQSSPRKGYTLAPSGTVNDFSDSGSEITSRSSTYSNS 1159
1287 --OPFVPRHGSTMAVYIPPLHMHAAHGFSTPSPGGV-----TSPATSAVAAWQ 1334
1160 FDSVPVSLHDERQHRHSVSIVETNIGMRERRTMIEPD-----QXSLGAYAPMSGRGL 1214
1335 CTPSP--SPCHRLRLAGGNTIIPRAIHERSHSTPAPRPLPSVDLSLES----- 1383
1215 YATATVYISFSTEELSODQDGRASLDADSGRSWTSKSSGSHDNIQTIOHRSWETLFP 1274
1384 -----SSVTT--FRDPLRKSVTSS--GSISSCDG-----YVHQOQ----- 1415
1275 GHTHPYSGDPALMAMSSSHMDQIMFSDHSTKYNRONQSESLQOASRAMASSTGYWG 1334
1416 -HYHLQY-----QQQOQOQNSSOHEPSPPVYTA----- 1442
1335 EDESGDTGTRKRGCKDVSILEASSLTSVTEETKPVMP-----AHIAVASTTGCLA 1390
1443 -----DORLLQOLSNNAVYTNL--NSPCQSTNTPPTPTPPNPOTATIQLSAPPTAAVM 1496
1391 RKEGRYRE-----PPPTPPGYIGIPITDPPEG--HSHPARRPPDYNAIORSRMV 1438
1497 HARSQHQOQLQOQOOSLAMPPEPPRY--NVP-----PLGSIVSH----- 1533
1439 ARSDDTAGPSSVQOPHCHPT-----SSRPVANKRO 1467
1534 --HQGTAGSRHLNMHGKTTGPOERWFPDRCPTTKQO 1568
RESULT 10
O43386
ID O43386 PRELIMINARY; PRT; 579 AA.
AC O43386;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE KIA001LB (Fragment).
GN KIA001LB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Connolly K.S., Gunning K.M., Davis C.A., Kadner K., Subramanian S.,
RA Miguel T., Lewis K.D., Fridlyand J., Alcivar D., Benke J.A.,
RA Bondoc M., Bowen E., Chiang A., Critz P., Jaklevic M.A., Lindo K.,
RA Lindquist K., Miller C., Patel S., Pisciotta C., Riley B.E., Rojeski H.,
RA Sarmiento R., Yu C., Montenegro M., Aerts A., Chung A., Abrajano A.,
RA Baker M., Gau C., Jett J., Ko C., Beall K., Woolley J.P., Stultz J.L.,
RA Kimerly W., Martin C.H.;
RT *Sequencing of human chromosome 5.*;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

QY	538	---	DKIRISLTPQKPYNDIGIGSODSIYGLAQTQKHIFIPALVYSGTSSSNDPLIOS	592
DB	744	SMDFTEAKISADLRP	-----SRSDPITS-	769
QY	593	HHRIIDF--SATPDPIDPVLRFKADDOOSRYTMSKDTJAKENVIOAIRFAVAT- PDQ		649
DB	770	---ISQYGPVNSCEPEHVKLTATKRDQTFKYLPLYKETSANQVVLALQEFMTGSGE		826
QY	650	YSLCEVYTPPEGVIKORRLPDOLSKLADRIOLSGRYLTKNNMETETLCSDEDAOELRES		709
DB	827	WSLCEVYTIIDGVIKORRLPQEMENLAERIALNSRYLTKNNSRSEPLVDELAPELKKA		886
QY	710	QISLQSLTEVEVATOLSMRNFELFNIEPEYITDLEFKIRSTSCANLKRFEVYNQETFE		769
DB	887	QTLSTLNAQVVAALDTLQDFSVFAIEPTFELMDLFLKDSKYSGSPKLEEPQLFNREXW		946
QY	770	IVASFLIETNQLKMKIIKHFIKIALNCRECKNNNSMFAITSGILNAPVALRTTWKTL		829
DB	947	WVAETICTERHQKRAKLKFKIKVAFRCYRDLRNTSMFAIMSGIDKPAVRRLHSSWERY		1006
QY	830	PKRYEKLFDQDLDDPDPSNNMAKYENVLNSQMLQEPPIEPYVIRKDLTFLHEGDSKVD		889
DB	1007	SKSYIRLMDIEHQVLDVPSRNMKRYOHLAEVAQEPVPVPIYVIKDLTFAMDGNATYSE		1066
QY	890	GLVNFELKLMIAKETIRHVGSMASVNMDDPALMERTKKRWSLGSLSQGSTNATVLDVAOT		949
DB	1067	KLINEKTRILIAKTSIRGVAKLSSASVYEIASMAE-----RSGVYVMDALLHNNSPENSIV		1120
QY	950	GGHKRVRRSSFLNKKLYEDAQMARVYQVYSNLELMDDESLOTLSIQCEPATNTLPK		1009
DB	1121	ATMRGSGMGKQONPKKAYEOLALMRKVSYLEGLHVDNDELMSYDIEPQVOTAHK		1180
QY	1010	--NPGDKPKVSETPSVAPRAGSOOKAOSLPPQOOPPPAKHI---NOGLQVPAYSLVPS		1064
DB	1181	GANSSTANIRRVSP-TPSSLSQSGAGADQSR---HHLELNGTC---SISAGC		1230
QY	1065	RKKVYVQDLPPEGICISPOLAKITLSLSESGSLERHKKAQEDTISNASSOLSSPTSPSS		1124
DB	1231	GSK-----FGVSPPAVQKMLSLVONSK-----YKGAPOPTSPSTARS		1271
QY	1125	PRKGYTLAPSGTVDNFSDSGHSEISRSSIYS		1156
DB	1272	LORNMPRV-TGRQATSSAOGPVOLNEETSTV		1302
RESULT 14				
QY	08VCKO	PRELIMINARY:	PRT:	528 AA.
AC	08VCKO:			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Similar to p02 domain containing guanine nucleotide exchange factor(GEF)1 (Fragment).			
OC	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=MAMMARY TUMOR;			
RL	Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: BC019702; AAH19702.1; -			
FT	NON TER			
SO	SEQUENCE	528 AA;	57244 MW;	64B0B78985BBC345 CRC64;
Query Match				
Best Local Similarity 45.8%; Pred. No. 6.7e-51;				
Matches 262; Conservative 69; Mismatches 166; Indels 55; Gaps 18;				
QY	939	TNAIVLDVAQTGGHKRRVRRSFLAKKLYEDAQMARVYQVYSNLELMDDESLOTLSL		998

Db	1	TNSNMDV--QGGAHKRRARRSSLLMAKKLYEDAOMARKVQYILSIDLIDTDEKFOHMSL	59
QY	999	QCEPPTINLPENPPDKKPVK-SETSPAPRAGSQQAOSLPQPOOPRPAHKINDGLQVP	1057
Db	60	QWEPARYGLTFLNLTFRKSAASSEKSPVPLRVSQOTAKVHLHOP-----HRYSOYLOVP	112
QY	1058	AVSLYPSRKVYPVKDLDPFEGINSPOALKLTLSSEGRSLERHKKQADTISNASSQLSP	1117
Db	113	AVNLHPRIKKGOAKD----HYLSTSLPQKGLCTPEEVSAYKH---TEDTISVASSLHSP	165
QY	1118	PTSPQSSPRKRYTTLAPGCTVDNFDSDSHSTISRSSSTYSNSSFSDYPSVSLHDERKORHSV	1177
Db	166	PASPNNSPRKGYTLTLPSSKCDNLSDSSHSEISSRSSISVNGSVDSMSAAGODERCSSHL	225
QY	1178	SIVENPILMGMEERTIIEPDQYSLGYAPMSGRGLYATATYISSPTEELSDQGDPA	1237
Db	226	AVPEPTGALETDTHPGCT--SDHSQILAHGMWLSKPCCLIKGVAVSSLSSEMSH---EIV	280
QY	1238	SLDADSGRGSWTSCSSGSHDNIOTIOHORSETL-PFGHTTFDYGDGAGLMASSSHD	1296
Db	281	VLEADSGRGSWTSCSSSHDNQSLONOKSMFLMSYRMHLD---DT---AEVEPTD	334
QY	1297	QIMPSDHSHTKRNQONSRESLEQAQSFASNASSTGYWGEDSECDGTIKRKGQDVSI	1355
Db	335	CEPCACPRKSGRTGQCKGSLFTWQLQSNWSSSL-SDTCFEPNYGTVAKRVLSEAPAA	393
QY	1357	ESSLSLSTVTEETKPVMPAHIVASTTGTLA-----RKGRYREPPPTPGYIGP	1410
Db	394	PDGLEPRTTD-----PYKYTTSSTDGKGLYCVTSFKKQDRKEPPPTPGILGIS	446
QY	1411	ITDPEEGSHAPKPPDYNALQSRWVARSDDTAGPSSVOOPHGHTPS---SRPVNKP	1467
Db	447	LADLEKG-PHPLTKPPDYSAVQSGKMWLNSLRPPA---PPSHTSAMVPSKIGSQDQ	502
QY	1468	WKKPESDPRLAPQYSGCFSTEDEDEQYSAV	1499
Db	503	RH----SHPKLADYADA--DSEADENQYSAV	528
RESULT 15			
08VCC8			
ID	08VCC8	PRELIMINARY; PRT; 876 AA.	
AC	08VCC8;		
DT	01-MAR-2002 (TREMblrel, 20, Created)		
DT	01-MAR-2002 (TREMblrel, 20, Last sequence update)		
DT	01-JUN-2002 (TREMblrel, 21, last annotation update)		
DE	Similar to CAMP-regulated guanine nucleotide exchange factor I (CAMP-GEF1).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	TISSUE=KIDNEY;		
RA	Strausberg R.;		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: BC020532; AAH20532.1; -		
DR	InterPro: IPR002373; CAMP_kin.		
DR	InterPro: IPR000595; CNMP_binding.		
DR	InterPro: IPR000651; RASGEFN.		
DR	InterPro: IPR001895; RASGRF_CDC25.		
DR	Pfam: PF000027; CNMP_binding; 1.		
DR	Pfam: PF00610; DEP; 1.		
DR	Pfam: PF00617; RASGEF; 1.		
DR	Pfam: PF00618; RASGEFN; 1.		
DR	PRINTS: PR00103; CAMPKINASE.		
DR	SMART: SM00100; CNMP; 1.		
DR	SMART: SM00049; DEP; 1.		
DR	SMART: SM00147; RASGEF; 1.		
DR	SMART: SM00229; RASGEFN; 1.		
DR	PROSITE: PS50042; CNMP_binding_3; 1.		

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2003, 17:42:44 ; Search time 41 Seconds

(Without alignments)
1516.415 Million cell updates/sec

Title: US-09-911-826A-2

Perfect score: 7721

Sequence: 1 MKPLAIPANHVGMOQGEKHS.....PYQSGGFSTEDDEDEQVSAY 1499

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	533.5	6.9	1234	1	YNX5_CAEEL
2	343	4.4	1595	1	SOS_DROME
3	325.5	4.2	1095	1	CC25_SACKL
4	314.5	4.1	1253	1	SC25_YEAST
5	309	4.0	1262	1	GNRP_MOUSE
6	307.5	4.0	1319	1	SOS1_MOUSE
7	303.5	3.9	1333	1	SOS1_HUMAN
8	303	3.9	1333	1	CC25_CANAL
9	290	3.8	1275	1	GNRP_HUMAN
10	288	3.7	1244	1	GNRP_RAT
11	282.5	3.7	1589	1	CC25_YEAST
12	277	3.6	911	1	STE6_SCHPO
13	245.5	3.2	1297	1	SOS2_MOUSE
14	244.5	3.2	768	1	RG12_HUMAN
15	242.5	3.1	1332	1	SOS2_HUMAN
16	238.5	3.1	768	1	RG12_MOUSE
17	227	3.1	2349	1	TPR_HUMAN
18	236.5	3.1	852	1	GNDS_MOUSE
19	234.5	3.0	914	1	GNDS_HUMAN
20	232	3.0	1435	1	LTRE1_YEAST
21	229.5	3.0	895	1	GNDS_RAT
22	229.5	3.0	1745	1	ZOI_MOUSE
23	218	2.8	1736	1	ZOI_HUMAN
24	217.5	2.8	777	1	RG12_HUMAN
25	211.5	2.7	778	1	RG12_MOUSE
26	211	2.7	1875	1	MLP1_YEAST
27	206.5	2.7	2845	1	APC_MOUSE
28	206.5	2.7	3122	1	DPOZ_MOUSE
29	200	2.6	1111	1	KIP1_YEAST
30	198.5	2.6	3210	1	KLEF_HUMAN
31	197	2.6	1066	1	KL61_DROME
32	196	2.5	2464	1	MABP_MOUSE
33	193.5	2.5	1953	1	BNT1_YEAST

34	191.5	2.5	1184	1	BIMC_EMENT
35	190.5	2.5	2459	1	MABP_RAT
36	190	2.5	2231	1	SEMI_YEAST
37	189	2.4	1462	1	NKCR_HUMAN
38	189	2.4	1727	1	ALML_SCHPO
39	187.5	2.4	2145	1	CYPA_PODAN
40	187	2.4	2472	1	NCR2_MOUSE
41	186.5	2.4	897	1	EP15_MOUSE
42	185.5	2.4	3924	1	ANK2_HUMAN
43	184.5	2.4	538	1	BUD5_YEAST
44	184.5	2.4	1957	1	YD86_SCHPO
45	184.5	2.4	2842	1	APC_RAT

ALIGNMENTS

RESULT 1
ID YN5_CAEEL STANDARD: PRT: 1234 AA.
AC P34578;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein T20G5.5 in chromosome III.
GN T20G5.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M., Smith A.;
RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 DEP DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CYCLIC NUCLEOTIDE-BINDING DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: Z30423; CAAB3013.2; -
CC PIR: S42368; S42368.
CC HSSP: P00515; 2BPk.
DR Wormpep: T20G5.5; CE23992.
DR InterPro: IPR000591; DEP.
DR InterPro: IPR000651; RasGEF.
DR InterPro: IPR001895; RasGEF_CDC25.
DR InterPro: IPR002373; CAMP_kin.
DR InterPro: IPR000595; CNMP_binding.
DR Pfam: PF00027; CNMP_binding; 2.
DR Pfam: PF00610; DEP; 1.
DR Pfam: PF00617; RasGEF; 1.
DR Pfam: PF00618; RasGEF; 1.
DR PRINTS: PR00103; CAMPKINASE.
DR SMART: SM00049; DEP; 1.
DR SMART: SM00229; RasGEF; 1.
DR SMART: SM00100; CNMP; 2.
DR PROSITE: PS00888; CNMP_BINDING_1; FALSE_NEG.
DR PROSITE: PS00889; CNMP_BINDING_2; FALSE_NEG.
DR PROSITE: PS50042; CNMP_BINDING_3; 2.
DR PROSITE: PS50186; DEP; 1.
KW Hypothetical protein; Repeat.
FT DOMAIN 206 336 CNMP-BINDING 1.

RA Prigozy T., Gonzales E., Broek D.;
RT "Identification and analysis of a DNA fragment from Saccharomyces
RT kluyveri that can complement the loss of CDC25 function in
RT Saccharomyces cerevisiae.";
RL Gene 117:67-72(1992).
CC
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,
CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M62964; AAA34479.1; -.
DR PIR; PC1114; PC1114.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGEF_CDC25.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 1.
DR PROSITE; PS00720; GDS_CDC25; 1.
KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;
KW Mitosis; Transmembrane.
FT NON_TER 1
FT TRANSMEM 959 980 POTENTIAL.
SQ SEQUENCE 1095 AA; 125605 MW; 7A8B3F90E31A44AC CRC64;

Query Match 4.2%; Score 325.5; DB 1; Length 1095;
Best Local Similarity 21.6%; Pred. No. 6.1e-08;
Matches 18; Conservative 110; Mismatches 240; Indels 225; Gaps 27;

QY 434 LEVNGQNFENIQ-LSKAMEILRNTHLSITVKTNL--FVEKELLRLSEKRRNG-----A 485
DB 380 LETSADTYKALNTYIDMIEVLEN--LDLTFYVLRNMSHDEKIDRSEELRKHAMTSLA 436
QY 486 PHLPKIGIDIKKASRYSPLDLA---DY-----EYVIGLEKNNKSKAN 525
DB 437 STIMEFEIQA---LHDIAIRQIMVOGLTLDDPEFCAIRDESCYCYDRDELSKYQ 492
QY 526 TVGGRNKLKILDKTRISIPOK-PYNDIGIGOSODSIYGLRQTKIIPALPVSGTLSS 584
DB 493 FL-----KQEFKAEELRNLLISQVEYNDLAFDLA--DQIVREAFKF---SDIGNTMLQ 542
QY 585 SNPDLSQSHRILDFSATPDLPDOVLRFKADOOS----- 619
DB 543 NVEGLIERENILNYARMMKSDLTALMKGEQKWFEDDFPMASSAEGENENLDFGDAQ 602
QY 620 -----RYMT-----SKDTAKVEVYQAIRE 640
DB 603 NKSEFRDIPWLDSEHYSILYDNKNGIKGTKEALEHLSHQSIDLSFNMLMLTFRS 662
QY 641 FAVT-----ATPOYSCEVSVPEGY-----IKORRLPDOL----- 672
DB 663 IFTTGEFLALVERYNL---YPEGVSYEYINWEEKKPKYKIRVYNIMKTLFSQYWT 718
QY 673 -----SKLADRIQLS-----GRYLLNNMET---ETLCSDEDAQEL 705
DB 719 PSYEPGLDWIGFAQAKSQKTSIGADYVLLSAIKGRISMGNLKNFPESINSDDOSSST 778
QY 706 -----LRESQISLLQSTVEVATQLSMRNFELFRNIEPEYID 743
DB 779 TTVPOSSRSSVAPVSSSTTGFRMRKLKLLDSDYAKQLTIKESLFLYKISPEFCLD 838
QY 744 DLF--KLRSKTSQANLKRFEVINOEFVWASLELRTNOLKMKIKIKIKIALHREC 801
DB 839 RTWGNKTCNNGSKNITEFISNSNHLTVVSEMIIVQOTDIKKRIQLIOPFINVAACHHEL 898

QY 802 KNFNSFAITISGLNLPVARLRTTWKLPKNKYKLEFODLDFPSRRNNAKYRNVINSON 861
DB 899 NNFSSILATISALYSSPIYRLKRTMAVAPREYKKLLETLMDSAKNFIRYQLLSIG 958
QY 862 LQPIILPFPVIKKDLTFLEHGDNS---KYDGLVNEK---LMIAKEIRHVR----- 909
DB 959 -DPPCVPEFGVYLSDLTFELFANGNDPLHRNTVYLVNFGKRYRLLEILKEISVQRSHYK 1017
QY 910 -----MASVMDPALMEFRTRKKRSLGSQSTNATVYLDVQGTGHHK 954
DB 1018 RYEDIAFISSLENLPSIRKOYAQSLRNPRTREVSTG---LNTSVNRYNNTKNGPENR 1074
QY 955 RV-RRSSFLNAKK 966
DB 1075 NTGKRLLKFGKAKK 1087

RESULT 4
ID SC25_YEAST STANDARD; PRT; 1253 AA.
AC P14771;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SCD25 protein.
GN SCD25 OR SDC25 OR YIL016W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M303;
RX MEDLINE=91094833; PubMed=1986220;
RA Damak F., Boy-Marcotte E., Je Roscouet D., Guilbaud R., Jacquet M.;
RT "SDC25, a CDC25-like gene which contains a RAS-activating domain and
RT is a dispensable gene of Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 11:202-212(1991).
RN [2]
RP SEQUENCE OF 668-1253 FROM N.A.
RC STRAIN=01136;
RX MEDLINE=89306677; PubMed=2545538;
RA Boy-Marcotte E., Damak F., Camonis J., Garreau H., Jacquet M.;
RT "The C-terminal part of a gene partially homologous to CDC 25 gene
RT suppresses the cdc25-5 mutation in Saccharomyces cerevisiae.";
RL Gene 77:21-30(1989).
RN [3]
RP FUNCTION.
RX MEDLINE=90260633; PubMed=2188363;
RA Crechet J.B., Pouillet P., Mistou M.-Y., Parmegiani A., Camonis J.,
RA Boy-Marcotte E., Damak F., Jacquet M.;
RT "Enhancement of the GDP-GTP exchange of RAS proteins by the carboxy-
RT terminal domain of SCD25.";
RL Science 248:866-868(1990).
RN [4]
RP FUNCTION.
RX MEDLINE=91156312; PubMed=2000228;
RA Rey I., Schweighofer F., Barlat I., Camonis J., Boy-Marcotte E.,
RA Guilbaud R., Jacquet M., Tocque B.;
RT "The COOH-domain of the product of the Saccharomyces cerevisiae SC25
RT gene elicits activation of p21-ras proteins in mammalian cells.";
RL Oncogene 6:347-349(1991).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -1- MISCELLANEOUS: SUPPRESSES THE CDC25-5 MUTATION IN YEAST (RESTORES
CC CAMP LEVEL) AND HAS SIMILAR FUNCTIONS AS CDC25.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entties requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M26647; AAA16565.1; -

DR PIR: S14177; S14177.

DR SGD: S0003939; SDC25.

DR InterPro: IPR000651; RasGEFN.

DR InterPro: IPR001895; RasGRF_CDC25.

DR InterPro: IPR001452; SH3.

DR Pfam: PF00018; SH3; 1.

DR Pfam: PF00617; RasGEF; 1.

DR Pfam: PF00618; RasGEFN; 1.

DR SMART: SM00147; RasGEF; 1.

DR SMART: SM00329; RasGEFN; 1.

DR SMART: SM00326; SH3; 1.

DR PROSITE: PS00720; GDS_CDC25; 1.

DR PROSITE: PS50002; SH3; 1.

KW Guanine-nucleotide releasing factor; Cell cycle; Cell division;

KW SH3 domain.

FT DOMAIN 26 98 SH3.

FT DOMAIN 74 79 POLY-ASN.

FT DOMAIN 434 437 POLY-ARG.

FT VARIANT 584 590 DVVAKFI -> V (IN STRAIN 0L136).

SO SEQUENCE 1253 AA; 144979 MW; 2DE2C9EC27E3E60D CRC64;

Query Match 4.1%; Score 314.5; DB 1; Length 1253;

Best Local Similarity 26.4%; Pred No. 2.4e-07;

Matches 106; Conservative 65; Mismatches 158; Indels 73; Gaps 12;

QY 680 QLSGRYLLKNNET-----ETLCSDDAQLRESQISLLQSTVEVATQLSMRNF 731

DB 909 EVNOKFKLGNIGEAFAPMKTLDOOLQDHYSGLYSTTE-SILAADPVLFAATQITLHE 967

QY 732 LRNIETPTIYIDLER---LRKTSKANLKREEVINQETFWVASILLETQNLKMKII 788

DB 968 IYCEITIFPCLOKIMKKNKYKTSYGASPGINEFISFANKLTFISVYKQKADSKRAKLL 1027

QY 789 KHFIRIALCRECKNNSMFAIISGLNAPVARLRTTWKELPKYKELKRODLDLPDSR 848

DB 1028 SHFIPTAECRKFNNSMTAIIISALYSPIRLEKTMQAVIYDPTDLQSLKLMPPK 1087

QY 849 NNAKYRNVLNSONLQPIIPLEPVYIKDLTLFHEGNSKVDGLVNEKRLMAKELRHVG 908

DB 1088 NFIVYNELKSLH-SAPCVDFPGVYLSDLTFDTSNPN---DYIVLEHGLKGVHDEKKYIN 1143

QY 909 RRAASVMDPALMFRIRKKRKRSLGSLSGSSTNATVLDVAQTGCHKRVRSSFLNAKKLY 968

DB 1144 -----FNKRSR-----LVDIIQ-----ETIYFKKTH 1164

QY 969 EDAQMARKVQYLSN-LELEMDESIQLTSLQCEPATNTLPKNPGDKPKYKSTSVAPAR 1027

DB 1165 YDPTKDRIVIECSNLENIPIHEKOYQLSLIEP-----KPKKK---VVPN 1208

QY 1028 AGSOCAKQSLPQPOQPPRAHKINQGLQVPAVSLYPSRKVP 1069

DB 1209 SNSNKSQKSRNDQDTDEKTKTKD-REPQGLHHTKKKAP 1249

RESULT 5

GNRP_MOUSE STANDARD: PRT: 1262 AA.

ID GNRP_MOUSE P27671;

AC 01-AUG-1992 (Rel. 23, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Guanine nucleotide releasing protein (GNRP) (Ras-specific nucleotide exchange factor CDC25) (CDC25mm).

DE RASGRF1 OR CDC25 OR GRF1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-BALB/c;

RC MEDLINE=931010996; PubMed=1396590;

RA Cen H., Lowy D.D.;

RT "Isolation of multiple mouse cDNAs with coding homology to Saccharomyces cerevisiae CDC25: identification of a region related to Bcr, Vav, Dbl and CDC24.";

RL EMBL J. 11:4007-4015(1992).

RN [2]

RP SEQUENCE OF 791-1262 FROM N.A.

RC STRAIN=Swiss; TISSUE=Brain;

RX MEDLINE=92289680; PubMed=1376246;

RA Martegani E., Vanoni M., Zippel R., Cocchetti P., Brambilla R., Ferrari C., Sturani E.P., Alberghina L.;

RT "Cloning by functional complementation of a mouse cDNA encoding a homologue of CDC25, a Saccharomyces cerevisiae RAS activator.";

RL EMBL J. 11:2151-2157(1992).

RN [3]

RP SEQUENCE OF 1031-1226 FROM N.A.

RX MEDLINE=92357779; PubMed=1379731;

RA Mei W., Mosteller R.D., Sanyal P., Gonzales E., McKinney D., Dasgupta C., Li P., Liu B.X., Brock D.;

RT "Identification of a mammalian gene structurally and functionally related to the CDC25 gene of Saccharomyces cerevisiae.";

RL Proc. Natl. Acad. Sci. U.S.A. 89:7100-7104(1992).

CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.

CC -1- TISSUE SPECIFICITY: BRAIN.

CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L20899; AAA02741.1; -

DR EMBL: X59868; CAA42525.1; -

DR PIR: S20730; S20730.

DR PIR: S22693; S22693.

DR MGD: MGI:98694; Rasgrf1.

DR InterPro: IPR001351; GDS_CDC24.

DR InterPro: IPR000048; IQ_region.

DR InterPro: IPR001849; PH.

DR InterPro: IPR000651; RasGEFN.

DR InterPro: IPR001895; RasGRF_CDC25.

DR InterPro: IPR000219; RhogEF.

DR Pfam: PF00169; PH; 2.

DR Pfam: PF00612; IQ; 1.

DR Pfam: PF00617; RasGEF; 1.

DR Pfam: PF00618; RasGEFN; 1.

DR Pfam: PF00621; RhogEF; 1.

DR SMART: SM00015; IQ; 1.

DR SMART: SM00233; PH; 2.

DR SMART: SM00147; RasGEF; 1.

DR SMART: SM00329; RasGEFN; 2.

DR SMART: SM00325; RhogEF; 1.

DR PROSITE: PS50010; DH_2; 1.

DR PROSITE: PS00741; DH_1; 1.

DR PROSITE: PS00720; GDS_CDC25; 1.

DR PROSITE: PS50096; IQ; 1.

DR PROSITE: PS50003; PH_DOMAIN; 2.

KW Guanine-nucleotide releasing factor; Repeat.

FT DOMAIN 22 130 IQ; 1.

FT DOMAIN 208 233 IQ; 1.

FT DOMAIN 244 430 DH; 1.

FT DOMAIN 460 588 PH 2.

FT DOMAIN 1025 1259 RAS-GEF.

FT CONFLICT 1033 1033 E -> D (IN REF. 3).


```

DR PROSITE; PSS0010; DH_2; 1.
DR PROSITE; PSS00720; GDS.CDC25; 1.
DR PROSITE; PSS0003; PH.DOMAIN; 1.
KM Guanine nucleotide releasing factor; 3D-structure.
FT DOMAIN 200 390 DH.
FT DOMAIN 444 548 PH.
FT DOMAIN 777 963 RAS-GEF.
FT DOMAIN 1244 1247 POLY-PRO.
SQ SEQUENCE 1319 AA; 150882 MW; 3286088A5BA0A46 CRC64;

Query Match
Best Local Similarity 19.6%; Score 307.5; DB 1; Length 1319;
Matches 224; Conservative 169; Mismatches 409; Indels 339; Gaps 47;

OY 236 NOVEKMKQVEEGER-VWVKEHRE--LDRTGRKHGIYIKGSEKTLNHLVEHSVDP 232
DB 240 NDVENFSTIVDHELSTVLILGHIEDTVEKTDGSPHPLVSGCFEDA-----EELAFDP 294
OY 293 --TFIEDFLLT--YRTFLSSPMVEGKLL-----EMFNDSPLRDKVTRVLLVNNHFNPD 343
DB 295 YESYARDILRPFHGHFLSOLSKPGALVLIQSIGEGFKE-AVQYVLPRLILAVYHCLNH 353
OY 344 FEEDPMATFLEFEENLREKKGHL-RLN-----IACAKAKRLMTITKESREA 395
DB 354 FE---LLKOLEEKSEDEQCECKKQAITALLVQSGMEKICSLAKRRRL-----400
OY 396 PLPFIILGSEKGFIFVDSGSKATEAGLKRGDILLEVNGQNFENIQLSKAM-EILR 454
DB 401 -----SESACRYSOOM-----KGOALAKMNEI-QKNIDWEGKDIGCCNEFTM 446
OY 455 NHTLSITVK--TNLFVFKELLRLSEKRNCAPIHL-----KIGDI 494
DB 447 EGIITVGAKKHEHIFILFPGLM--ICCKSNHGPRLRGASSAFYRLKEKFFMKVQINDK 504
OY 495 KKAASR--SIPDLAVDEOYI-----GLEKVNKKSAFTVGGNNKLKILDKTRISILPQ 547
DB 505 DDTSEYKHAFAEILLKDGNSYIFSAKSAEEKNNMALISIQYRSTLERMD-----555
OY 548 KPYNDIGIGSQSDSIVGRQTHIPTALPVSGTLESSNPDLLQSHHRL-DRSATPDLR 606
DB 556 -----VTVLOEKEEEDMRLPASVYRPAED--SEENLFEENVOPKAG 597
OY 607 DQYLK--VFKAQOQSRVIMISKDTAKEVYIQAIREFAVTATPDQYSLCEVSVTEGVI 663
DB 598 IPIKAGTVLKLIERLTHMYADP-----NFVRTPLTT--YRSCRPQELSLILI 645
OY 664 KORRLDQLSKADRIQLSG-----RYYLKNMTEFLCS-----D 699
DB 646 ERREIPEPEPTEDRIAENGDOPLSAELKRFKREYIOPVQLVNLVCRHWVHNFEYDE 705
OY 700 EDQOELLRESQ-----710
DB 706 RQADLLQRHEEFGIVRGKAMKMWESTIKIIOKKIARDNGCHNITROSSPTPYEMHI 765
OY 711 -----ISLQLSVEVATQLSMNRFELRNIEPTVEYIDLEFLRSK-TSCANLKREE 761
DB 766 SRGHETFDLTLHPILIELARQLTLESDDLXRAVQPSSELGVSWTEDKEINPNLLKMI 825
OY 762 EYVNOCTFWASITLRETQOLKMKLIIFHFKIKALHCRCKNFNMPAIIISGLNLPVAR 821
DB 826 RHTTNLTLEFEKCIETENLEERVAVVSRIETILQVFOELINNGVLEVASNNSSPYVR 885
OY 822 LRTTWKELNKYEKLFQDLODLFDP--SRNMAYRVNVLNSQNQPPIILPFIYIKKDLTF 879
DB 886 LDHTEFQIOPSROKILIEEHLSESDHYKYTLARLSI-----NPPCVPFEGYLTNLIK 939
OY 880 LHEGNSKVDGLVFEKLEKIAEIRHVGKMASVNMDDPALMFTKRRKMSLGSLSQSGST 939
DB 940 TREGNP-----EVLRRHGKEL-----INFSKRRVAELITGEIOQYON 976
OY 940 NATVLDVADQGTGHHKRVRRSSFLNAKKLYEDA-----QMAKKVQOYLINLELMEDESIO 994
DB 977 QPYCLRVER-----DIKRFFENILPMGNSMEKEFTDYLFNKSLEI-----1016

```

```

OY 995 TLISLQCEPA-TNTPKPNPGDKK---PVKSE-TPSPVAPRAGSOOKAOSLPQOQPPPAHK 1049
DB 1017 -----EPHHPKPLPRP--KISTYPLKSPGVPSNPRECTMHH-----PPLQOEP-----1060
OY 1050 INQGLVPVAVSLYPSRRKVPVKDLPPEGINSQALKKILSLSESGSLERHKKQAEPTISN 1109
DB 1061 -----RKISYSRIIP-----ESETESTASA 1079
OY 1110 ASS---QLSSPPTSPQSSPPKGYTLAPSGTVDNFSDSGHSEISRSISVNSSFDSVPVS 1166
DB 1080 PMSRPPLTPPPASGTSNTDVCVSFDS--DHSASPFHRSASVSISLSKGTDEVPVP 1136
OY 1167 LHDERQORSVSIVETN---LGMGRMERRTMIPDDQYSLGSVPMSEGRGLVTFATVYISS 1223
DB 1137 PPVPPRRRPPESAPAESPSKINSKHLDSPALPPROPTSKAISP---RISISDRTISID 1192
OY 1224 P 1224
DB 1193 P 1193

RESULT 7
SOS1_HUMAN
ID SOS1_HUMAN STANDARD; PRT; 1333 AA.
AC 007889;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Son of sevenless protein homolog 1 (SOS-1).
GN SOS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93262494; PubMed=8493579;
RA Charlin P., Camonis J.H., Gale N.W., van Aelst L., Wigler M.H.,
RA Bar-Sagi D.;
RT "Human Sos1: a guanine nucleotide exchange factor for Ras that binds
RT to GRB2.";
RL Science 260:1338-1343(1993).
RN [2]
RP STRUCTURE BY NMR OF 422-551.
RX MEDLINE=98043737; PubMed=9374522;
RA Zheng J., Chen R.H., Cordlan-Garcia S., Cahill S.M., Bar-Sagi D.,
RA Cowburn D.;
RT "The solution structure of the pleckstrin homology domain of human
RT SOS1. A possible structural role for the sequential association of
RT diffuse B cell lymphoma and pleckstrin homology domains.";
RL J. Biol. Chem. 272:30340-30344(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 198-551.
RX MEDLINE=99005193; PubMed=9790532;
RA Soisson S.M., Mimmual A.S., Uy M., Bar-Sagi D., Kuriyan J.;
RT "Crystal structure of the Dbl and pleckstrin homology domains from
RT the human Son of sevenless protein.";
RL Cell 95:259-268(1998).
CC 1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC 1- SUBUNIT: INTERACTS WITH GRB2.
CC 1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC 1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC 1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```
CC EMBL: L13857; AAA35913.1; -
DR PDB: 1AME; 25-FEB-98.
DR PDB: 1DBH; 23-DEC-98.
DR Genew: HGNC:11187; SOS1.
DR MIM: 182530; -
DR InterPro: IPR001331; GDS_CDC24.
DR InterPro: IPR002119; Histone_H2A.
DR InterPro: IPR004822; Histone_core.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGEFN_CDC25.
DR InterPro: IPR000219; RhogEF.
DR Pfam: PF00169; PH. 1.
DR Pfam: PF00617; RasGEFN. 1.
DR Pfam: PF00618; RasGEFN. 1.
DR Pfam: PF00621; RhogEF. 1.
DR SMART: SM00414; H2A. 1.
DR SMART: SM00233; PH. 1.
DR SMART: SM00147; RasGEF. 1.
DR SMART: SM00229; RasGEFN. 1.
DR SMART: SM00325; RhogEF. 1.
DR PROSITE: PS00741; DH. 1; FALSE_NEG.
DR PROSITE: PS50010; DH. 2; 1.
DR PROSITE: PS50020; GDS_CDC25. 1.
DR PROSITE: PS50003; PH_DOMAIN. 1.
DR Gnanine-nucleotide releasing factor: 3D-structure.
KW DOMAIN 200 390 DH.
FT DOMAIN 444 548 PH.
FT DOMAIN 777 963 PH.
FT DOMAIN 1258 1261 POLY-PRO.
SQ SEQUENCE 1333 AA; 152463 MM; CGB99C0A11A8DE45 CMC64;

Query Match 3.9%; Score 303.5; DB 1; Length 1333;
Best Local Similarity 19.0%; Pred. No. 8.6e-07;
Matches 221; Conservative 175; Mismatches 397; Indels 371; Gaps 47;

QY 236 NOYEKNQKVEEGEI-VVMKREHE-LDRTGRKHIVYKGTSERLTMLVEE----- 286
DB 240 NDVENIFSRVLDIHELSEKLLGHIEDTVEMTDESSPHVLGSCFEDIAELAEDEPYEVA 299
QY 287 HSYVDPTFIEDFLTYRTFLSP-----MEVGKLEMFENDPSLDKVTRVVLLAVNNH 340
DB 300 RDLRPGFHRL-----SQLSKPGALYLOSIGGEFKE-----AVGYVLRLLLPYHC 350
QY 341 FNDEGDPAMTRLEEFENNLERKMGHL-RLN-----IACAARRLMTLTKPS 392
DB 351 LHFFE--LLKOLEKSEDEDEKCLKQATLALNVQSGMEKICSKLARKRL----- 400
QY 393 REAPLRPILLGSGKRGIVDSVDSKATFAGLKRGDOLLENGGNFENIOLSKAM-E 451
DB 401 -----SESACRYSQM-----KKGOLAIRMNEL-QKNIDGMEGKDIGQCCNE 443
QY 452 ILRNNHLSITVK--TNLFYFKELTLRLSEKRNAGALP-----KI 491
DB 444 FIMEGTLTRGAKHEHRIPLFDGLM--ICKSNHGQRLDGASNAEYRLKEKFEKRVQI 501
QY 492 GDIKKASRY--STPDLAVDVQVY-----GLEKYNKKSKANTYGGRRKKLILDKTISI 544
DB 502 NDDADNTNEYHAPEIILKDNSVYFSKASABEKNMMAALISLOYRSTLERMLDVTM--- 558
QY 545 LPQKPYNDIGISQODSIVGLRQTKHIPTALPVSGTLSSNPDLQSHRILDFESATPD 604
DB 559 -----LQEEKQOMRLPSADYVRRAEPP--SEENIIFENNQPK 595
QY 605 LPDQVLR--VFRAQQSRYIMISKDTTAAEVYIOAIRFAVATATPDQYSLCEVSVPEG 661
DB 596 AGPIIAGVIVIKIELRTYHMYADP-----NRYRTFLT--YRSPCKQQLSL 643
QY 662 VIKORRLPDQLSKIADYIQLSG-----RK----- 685
DB 644 IIEFEIPEPEPTADRIAIENGQDPLSAELKRFREYIQPVOLRVLVNVCRRHWEHHFYD 703
```

```
QY 686 -----YLNKME-----TETLCSDPEDAQELLRES----- 709
DB 704 FERDAVLLQMEEFICTYVRCKANKKWEESTYTIQKRIKARDGPGHNITFQSSPPTVEW 763
QY 710 -----QISLQSTVEVATQLSMRNPELFRNIETPEYIDDLFKLSRK--TSCANLKR 759
DB 764 HIRPGHIEFTDLTLHPRIEIAQLTLLESDDLRAVQPSLVSQVTKDEKELNSNLLK 823
QY 760 FEEVINOETFWASEILRETNOIKRMKIKHFKIALHCRCKENFSMFAISGLNLAVP 819
DB 824 MIHNTNLTLMFEKCIYEVENLEERVAVVSRIEIILOVFOELINFGVLEVVSAMNSSEV 883
QY 820 ARLRTWEEKLPNKYEKLPQLODLFDP--SRNAKRRNVILNSGNLOPPIIPFPVKKDL 877
DB 884 YRLDHTPEQIPSRQKKLLEAHNELSDHYKKYLAKLRSI-----NPPCVFPGIYLTNI 937
QY 878 TFLHEGNDKSVQDLVNFYEKLRIAHVGRMASVYMDPALMFTRTKKRWSLSLSOG 937
DB 938 LKTEEGNP-----EVLKRGKEL-----INFSKRRKVAELTGEIQQY 974
QY 938 STNATYLDVAQGTGHHKRRVRSFLNAKKLYEDA-----OMARKVQYLSNLELMEDES 992
DB 975 QNOPYCLRVES-----DIKREFENLNPNGSMEMEKEFTDYLFNKSLEIEPRN 1020
QY 993 LQTLSDCEPATNTLTPKNPGDK--PVKSE-TSPVAPRAGSQQKASLPQPOQOPPAH 1048
DB 1021 -----PKPLRFP--KKYSTPLASPGVPSNPRGTRH-----PTPLQOEP--- 1060
QY 1049 KINQGLQVPAVSLYPSRKRVAVKDLPPFGINSPOALKLILSEEGSLERHKKQAEDTIS 1108
DB 1061 -----RKISYSRIP-----ESETESTAS 1078
QY 1109 NASS---QLSSPTSPQSSSRKGYTLAPSGTVNFS--DSGHS----- 1146
DB 1079 APNSPRTPLTPPPASGASS-----TTDVCASFPSDHSPPHSSNDVFIQVTLP 1127
QY 1147 ---EISRSISIVNSSPDSVPVSLHDERORHSVSIETN---LGGRMRRTMIEPDQY 1200
DB 1128 HGRSASVSSISLTGKIDEVPPVPPRRPPRPSAPRESSPKIMSKHLDSPAIIPROP 1187
QY 1201 SLGSYAFMSBEGRLYATATVYISSP 1224
DB 1188 TSKAYSP-----RYSISDRTSISDP 1207

RESULT 8
CC25_CANAL
ID CC25_CANAL STANDARD; PRT; 1333 AA.
AC P43069;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell division control protein 25.
GN CDC25 OR Cdc25.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=792-1;
RX MEDLINE=93238685; PubMed=8477693;
RA Goldberg D., Marbach T., Gross E., Levitzki A., Simchen G.;
RT "A Candida albicans homolog of Cdc25 is functional in Saccharomyces cerevisiae.";
RL Eur. J. Biochem. 213:195-204(1993).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GTP BY GTP. THIS
CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,
CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M9160; AAA34329.2; -
DR HSSP: P29354; 1GFC.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGRF_CDC25.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00618; SH3; 1.
DR Pfam: PF00617; RasGEFN; 1.
DR Pfam: PF00618; RasGEFN; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3; 1.
DR SMART: SM00147; RasGEF; 1.
DR SMART: SM00329; RasGEFN; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00020; GDS_CDC25; 1.
DR PROSITE: PS00002; SH3; 1.
KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;
KW Mitosis; SH3 domain.
FT DOMAIN 30 94 SH3.
SQ SEQUENCE 1333 AA; 152012 MW; DAFB90D6AC192483 CRC64;

Query Match 3.9%; Score 303; DB 1; Length 1333;
Best Local Similarity 19.5%; Pred. No. 9,1e-07;
Matches 214; Conservative 157; Mismatches 339; Indels 390; Gaps 44;

QY 23 ADFTKHLVDSLPHOYTHVS-----SSHGCGSTTSDGSSSLSDIYQAT 66
DB 432 ASFAQIYINGILSLVHLSYLEGFNEGRLFGYDMGKLNKSSSSARQSPASSLSTIRGCS 491
QY 67 E-----SEAGMDLSGLPETAVDSDDEDEIE-----RASDPLMSR 104
DB 492 DSDTRFAOKLSQDRNSEGNDMNYINOLVY--FIDNLRENVNSIVKFLKLSANKKIKNS 549
QY 105 DIYVDCLEKPRIDRTDDDIQILEFMHQL-PAF-----ANMTASVRELCAVMVF 133
DB 550 DY-----DSSASDDEGDRFDILPQVYPRFLVDFENGMCNPFSTKNTVLANVSGD 602
QY 154 AVERAGTIVLNGEELDSVILNGSVETVYDPKAEILCMGNSFCVPTMDKEVMKV 213
DB 603 DLNKRHTKIIIDHSAYDLS-----QIVDKI 629
QY 214 MTKVDDCOFVCIADODYCRILNOVEKNMKEVEGEIYVAKKEHRELDRTGRKGTIVIK 273
DB 630 -----VQACE-----NILEALDPKVQNTFYNNEMLNERNTQI----- 662
QY 274 GTSERLT-MHLVDEHSVYDPTFIEDLLIYRTFLSSPMVEYKKL---LEWFPNDPSLRDKV 329
DB 663 ---LRLTYKSLVYCSAMVD--LIESF--DFTVFCSVYRHGNADIDEDSEYDPS----- 710
QY 330 TRVVLVNVNHFDFESD-PAMTRLEEFENNLEREMGSHLLNLTACAKKRRLMTL 388
DB 711 ----WGDHYDSNLSFDYVVLLEPF-----RLKQELHDLVAKIKMTQS---LTL 753
QY 389 TKPSREAPLFFILLGSEKGFIVDSVDSGSKATEAGLRGQDILLEVNOQNFENIOLSK 448
DB 754 EDPE-----VF-----KGLKEEDPL-----EYNRISK 776
QY 449 -----AMELIRNN-----THLSITVKNLFEVKEELLTRLS 478
DB 777 IPKEKALLSLTIKEQLSFKDGAISLNPDTLISGLVLAIAKTKVLLITOOOL---- 832
QY 479 EEKRNAGPHLPKIGDIKKARSYSPDLAVDEQYIGLEKYNKKSKANTVGGRRKLLKIID 538
DB 833 EEEETIINATRV---MDQNFQOLLVBRNNTSSSEKADNS--YVVGSHKK----- 880
QY 539 KTRISILPQRPYNDIGIGOSODSIYGLROTKHIFPALPVSGTLSSSNPDLLOS---HHR 595

DB 881 -----STDVPMYLEGDEDEYELL-----LDVKGNIKGSKKALVSHLTHHL 920
QY 596 ILD--FSATPDLDDQVLRVKAKDQGSRYIMISKDTAKEVQIAIRFATATAPPDQYSLC 653
DB 921 SLDSNFNA-----VF-----LWFSMSMSIGELISLILARENTE----- 954
QY 654 EVSVTEGV-----IKORRLPDQLSKL-----A 676
DB 955 -----PEGLSLYEENLWMSKRRNPRIKILKLEKMWMSYNEPVLRRNLTFRAHS 1010
QY 677 DRIQLSGR-----YLLKNMETETLCSDED-----AOELLRESOIS---LLQLSTV 719
DB 1011 DOVQVTSIGMLVNYLERLLRGERIYVERDPVJPTKPAFLFTGSSLSKKRPVMDIDY 1070
QY 720 EVATQLSMRNPELPRNPEPEYIDDLFLKLS--KTSQANLKRPEEVNOCETFWASILR 777
DB 1071 ELARQLTRREFKLYCKTKTKACLAKWGGKSGSESIDSTIOPFKASNOILNFEYGMILR 1130
QY 778 ETNOLKRMKIKHFIKLTALCRECKNPNFSAITSGNLAPVARLPTWEKLPKYEKLF 837
DB 1131 KADPKRRVQITIRFYQVADKCRQYNNFSSMTAITLSALYSPIRHLKKTWEYMADALSNL 1190
QY 838 QDIQDLFDPSPRNMAKYRNVLNSONLQPIILPFPYIKKDLFTLHEGNSKV--DGLVNF 894
DB 1191 KNNMKLNSSRNFEYRDVLFKGSE-PCVPFGVYLSDDLTFVYHGNDPYLYNRTQGVNF 1249
QY 895 EKLMTAKETIRHYGR-----MASVYNDPALMFR 922
DB 1250 AKRAKTSVLSIGDIFRFTTGYNFQVEPELQFLDAPWEKCPITDEQYQISLNEP----- 1304
QY 923 TRKKWRSLGSLSGSTNAT 942
DB 1305 ---REOPAGASNSNSTNAT 1321

RESULT 9
GNRP HUMAN
ID GNRP HUMAN STANDARD; PRT; 1275 AA.
AC Q13972;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Guanine nucleotide releasing protein (GNRP) (Ras-specific nucleotide
DE exchange factor CDC25).
GN RASGRF1 OR CDC25.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95129875; PubMed=7828890;
RA Wei W., Das B., Park W., Broek D.;
RT Cloning and analysis of human cDNAs encoding a 140-kDa brain guanine
RT nucleotide-exchange factor, Cdc25GEF, which regulates the function of
RT Ras.
RL Gene 151:279-284(1994).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IO DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L26584; AAA58417.1; -
DR Genew; HGNC:9875; RASGRF1.


```

DR InterPro: IPR001331; GDS_CDC24.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000651; RasGEF.
DR InterPro: IPR001895; RasGEF_CDC25.
DR InterPro: IPR000219; RhoGEF.
DR Pfam: PF00169; PH; 2.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF00617; RasGEF; 1.
DR Pfam: PF00618; RasGEF; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00233; PH; 2.
DR SMART: SM00147; RasGEF; 1.
DR SMART: SM00229; RasGEF; 1.
DR SMART: SM00325; RhoGEF; 1.
DR PROSITE: PS50010; DH_2; 1.
DR PROSITE: PS00741; DH_1; 1.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PROSITE: PS50096; IQ; 1.
DR PROSITE: PS50003; PH_DOMAIN; 2.
KW Guanine-nucleotide releasing factor; Repeat.
FT DOMAIN 22 129
FT DOMAIN 204 229
FT DOMAIN 240 426
FT DOMAIN 467 584
FT DOMAIN 1038 1272
SQ SEQUENCE 1275 AA; 145381 MW; 86C6F54AA1E451F1 CRC64;

```

Query Match

```

Best Local Similarity 19.4%; Score 290; DB 1; Length 1275;
Matches 218; Conservative 162; Mismatches 393; Indels 348; Gaps 48;

```

```

QY 3 PLAPAHGWAGQ-----EKSLPADFTKL-HLTDLSLHQVTVHSSH-----S 46
DB 386 PRYLTLHDVLAHPHHEVERNSLIDYAKSKLELSRIMHDEVTENIRKNALEIRMIIE 445
QY 47 GCSTISDSSSSSL--SDIYQATESEAGDMDSGLPETAVDSEDDDEEDIERASDPLMR 104
DB 446 GCELLDTSOIFVAGSGSLIYQPNSEKGIKTRGRIGSLSLREG-----ERQGF-LFSK 497
QY 105 DIV-----RDCKLEKPIDRTD-----DIQL-----LEFMHQ 132
DB 498 HLICTRSGGKHLHTKNGVLSLIDCTLEPESTEEBAKSGODIDHLFFKIVGEPRDS 557
QY 133 LPAAANTMSVRRELCAVMFAVVERACTIYVNGCEELSDSVYVTEVETYPDGAELI 192
DB 558 PP-----FTVI-----LVASSRQEKAAWTSISOCVD-----NI 586
QY 193 LCMG-----NSFGVSPTMDKEVMKGVMTKY--DDQCFVCI AQODYCRILNQVEKNM 242
DB 587 RCNGLMMAFEENSKVYVPMIKRTRETRAEKSRSDASLYC-----DDVIDRF 636
QY 243 QKVEEEGIYVWKHRELDRGTGRKHIVIKGTSERLTMHVEHSVVDPTFIDFLITY 302
DB 637 SKTMNSCKVLDIR-----YASVRELLERLT-----DLRFLSIDFLMTFLHSY 678
QY 303 RTFLSSPNEVCKLLEMPNDPSLDKVRVYLLWV-----NNHNFDEFGDPAMT 351
DB 679 RVP--TTAIVLADKLITTYIKRP-----ISATPARMLRSLIELLPASGQNNKL--LYGEPRKS 730
QY 352 -RFLFEEN--NLEREKNGHRLNLINACAAKAKRRLTTLTKPSREAPLPTLLGG---- 404
DB 731 PRATRKFPSPPLSTTKTSSPSR-----RRKLISLNIIP-----ITGSKALD 772
QY 405 -----SEKFGITFVDSVDSGSKATEAGLRGQDILEVNGQNFENIOLSKAMELRNNT 457
DB 773 LAGSLSCNSNGYTSWYSAMSPFSKAT-----LDTSK----- 803
QY 458 HLLSTIVKNLVFPKELLRLSEKRRNGAPHLPKIGDI--KKASRSYIPDLAVDVEQVIGL 515
DB 804 -----LYVSSSTFNKIPDE--GDITPERKPEPSALSOSSEVSMREES--DIDQ----- 848

```

```

QY 516 EKVKKSKANTYGVGRNK---LKKILDKTRISILPOKPYNDIGISQSDSVIGLROTKHI 572
DB 849 ---NQSDGDDTSPKSPPTTPKSVKNKNSSEPLFYSYNN--GVYMTSCRELDNNRSALSA 904
QY 573 PVALPVSGTSLSSNPDLDSHHRILDFSATPDLPOVLRFVKAQDS--RYIMISKDTTA 630
DB 905 ASAFATATGANGCTPNKREKRYRMSIASAG-----FPDDONGKEFVIRRAATN 954
QY 631 KEVYQAIREFAVTATPD---QYSLCEVSVTPPEGYIKORRLPDQLSKLADRLQLSGRY 686
DB 955 R--VLNVLRHWYSKHSQDETDELCKVYIGFLFVYMHDPPELLTOERKAANI----- 1005
QY 687 LKNMTEPLCSDEDAQELLRESQISLQL-----STVEVATQLSMRNF 731
DB 1006 -----IRTLTQEDPG-----DNQITLEETIOMAEGVKAPEFNHSALEIAQLTLLDHL 1054
QY 732 LFNINIEPTETI--DQLFKL---SKTSCANLKRFEVINOETFWASSELRETNOUKRK 786
DB 1055 VFKKIYEEFFGGQWMLKEKNERTPYIMKTKHFNDISN---LASELIRREDINARVS 1110
QY 787 IIRKFIKIALHCRECKNPNMFALISGLNLAPYARLRTTWKLPNKYEKLFODLDPDP 846
DB 1111 AIEKWAADVADICCLHNHYNAVLEITSSMNRSAIFRLKTKWLYSKOTKALIDKLOLYSS 1170
QY 847 SRMAKYRNVLSNQLOPPIPLPYIKKDLTFLEHGNSKV--DGLVNEEKLRLMAKEIR 905
DB 1171 EGFRKNLREAL--KNDDPCVPYLGMYLTDLARIEGTPRYTEGDLVNSKRMISHIIR 1228
QY 906 HVGMAVSVMNDPRLMPTTRKKKWSISLSQSGSTNATVLDVNOTGHHKRRVRRSSFLNAK 965
DB 1229 EI-----ROFOQTAY 1238
QY 966 KLYEDAQMARKVOYLSNLEMEDESLQTLISQCEPATNT 1006
DB 1239 KIEHQA---KTYQYLLDOSFVWDESLVLESSLRIEPKLPT 1275

```

RESULT 10

ID	GNRP_RAT	STANDARD;	PRT;	1244 AA.
AC	P28818;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Guanine nucleotide releasing protein (GNRP) (P140 Ras-GRF).			
GN	RASGRF1.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=92350260; PubMed=1379346;			
RA	Shou C.; Farnsworth C.L.; Neel B.G.; Feig L.A.;			
RT	"Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor for Ras p21.";			
RL	Nature 358:351-354(1992).			
CC	- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.			
CC	- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.			
CC	- SIMILARITY: CONTAINS 1 PH DOMAINS.			
CC	- SIMILARITY: CONTAINS 1 IQ DOMAIN.			
CC	- SIMILARITY: CONTAINS 1 IQ DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; X67241; CAA47666.1; -.			

DR PIR: S29083; S29083.
 DR InterPro: IPR001331; GDS_CDC24.
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR000651; RasGEF.
 DR InterPro: IPR001895; RasGRF_CDC25.
 DR InterPro: IPR000219; RhogEF.
 DR Pfam: PF00169; PH; 2.
 DR Pfam: PF00612; IQ; 1.
 DR Pfam: PF00617; RasGEF; 1.
 DR Pfam: PF00618; RasGEF; 1.
 DR Pfam: PF00621; RhogEF; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00233; RasGEF; 1.
 DR SMART: SM00147; RasGEF; 1.
 DR SMART: SM00229; RasGEF; 2.
 DR SMART: SM00325; RhogEF; 1.
 DR PROSITE: PS0010; DH; 2; 1.
 DR PROSITE: PS00741; DH; 1; 1.
 DR PROSITE: PS00720; GDS_CDC25; 1.
 DR PROSITE: PS00096; IQ; 1.
 DR PROSITE: PS0003; PH_DOMAIN; 2.
 DR PROSITE: PS0003; PH_DOMAIN; 2.
 KW Guanine-nucleotide releasing factor; Repeat.
 FT DOMAIN 22 129 PH 1.
 FT DOMAIN 204 229 IQ.
 FT DOMAIN 240 426 DH.
 FT DOMAIN 456 582 PH 2.
 FT DOMAIN 1007 1241 RAS-GEF.
 SQ SEQUENCE 1244 AA: 142666 MW: 48647879E842AF6B CRC64;

Query Match Best local similarity 3.7%; Score 288; DB 1; Length 1244;
 Matches 201; Conservative 111; Mismatches 368; Indels 350; Gaps 42;

DR 17 EKHSLPADFTKL-HLDSLHPQYTHVSSH-----SGCSTSDSGSSSL--SDIY 63
 DR ERNSLDYAKSKLELSEVHMDSETEINIKNLAIERMITECEILDITQTFVROGSLI 464
 DR 405 QATESEAGDMDLSGLPETAVIDSEDD-----DDEEDIERADPLMSRIYR 108
 DR 465 QVPMSEKGIKINKRGLSGLSKKEGROCFLESHLLICTRGSSSKLLTNKGVIS--LI 521
 DR 109 DC-LEKPDIDRTDDIEQLLEFM-----HOLPAFANTMGRVRELCAVNFAYVE-- 157
 DR 522 DCLLLDPEKMDDDKQGEVDHDFKIWEPRKSPPTTYLLVASSKREKAAVSDIICV 581
 DR 158 ---RAGTIVLNDGEELDSVSLNGSVETYPD--GKAEILCMGNSFGVSPMDKEYM 211
 DR 582 DNIRCMGIMNAFEE-----NSKYTVPMQIKSDASLYCDDVDIRSKTMNS---- 627
 DR 212 GVARIVDCQFCIAQDYCRLLNOYERKMQVEEGELVMYKREHRELDRTGTRGHIY 271
 DR 628 -----CKVL-QIR-----YAS 637
 DR 272 IKGTSERLTHLVEHSHVDPFTIEDFLTYRTFLSSPMVEGKLLMEWNPDLRQVTR 331
 DR 638 VERLLELT-----DLRFLSIDFLNTPFLHYRF-IDAVVLDKLISYIKKPTALPARS 691
 DR 332 VILLWVNNH-----FNDPEGDPAMTRFLEEFENLEREKMGHRLNLINCAAKAKRIM 386
 DR 692 LEILFSSSHNTKLLYGDAKSPRASR---KSSPPP-----LAIGTSSPVRKRKL 728
 DR 387 TLKPSSEALPFLILGSGSKGFIYDVSOSKATEAGLRGDDQLLEWNGNFENIQL 446
 DR 739 SLNIP-----ITGG-----748
 DR 447 SKAMEILR-----NNTHLST-----TVKTNLEVFKEELTRLSEKRNCAPIPKIG 492
 DR 749 -KALELASLGCPSDGYTNHISPSPECKTLDTSKLCVASSLRTPE-----IDMT 799
 DR 493 DIKASRYSTPDLAVDEVOYIGLEKKVKSANTVGGRNKLKLLDKTRISILPQKPYND 552
 DR 800 TLESSSGFRP-----TSDILKEESDDQSD-----VDDTEVSPTPKSFRN 841

QY 553 IGIGOSODDSIVGLRQTKHIPALPVSGTLSSSNPDLOSRRHRLDPSA----- 601
 DB 842 -----RITQEPFLFNYSGLIMTCR-DLMSNRKPLSATSAFALATAGANE 886
 QY 602 TPDLPOVLNV-----FKADQOS--RYIMSKDTAKEVYQAIRFEVATAPD----- 648
 DB 887 SPANKELIYRMSLANGYSDQGNIDKEFYIRRAATNR--VLNVLRWVYKHQDEFD 944
 QY 649 --QYS-LC---EVSYPEGVYKORRLPDQLSKLADRIQLSGRYLYKNNMETFLCSD 702
 DB 945 LKRYVICFLEEVYHDPDLIPQERKAAANIMRTLOEITEHN-----SMDELLMTEGV 1000
 QY 703 QELLRSQISLLDLSITYEVATQLSMRFELEFRNIEPEYIDDLFKLSKTS-----CANL 757
 DB 1001 KTEPFENH-----SAMEIAEQLTLLDHLVEKSIPEEFGQGMADKMERPTPIYKTT 1054
 QY 758 KREEVINQETFWVASEILEFTNQLKRMKTIKRFIKALCRECKNPSMEAIISGLNLA 817
 DB 1055 RHFNNHISN-----LIASEILLNEVSARASTIEKRVAVADICRCLHYNNAVLEITSSINRS 1110
 QY 818 PVARLTWEKLPNKYEKLEFQDLDLEDPSPRNMAKRYRVNLSONLOPPIPLPEVYIKDL 877
 DB 1111 AIPRLKKTWLVKSGKQFSLPDKLQKLVSSDGRFKNLRETL--RNCDPVCVPLGMVLTDL 1168
 QY 878 TELHEGNDKVV-DGLVNFELRLMAKEIRIVGRMASVNMDPALMFRKKKNSLSLSQ 936
 DB 1169 AFLKEGPNYTEDGLVNFESKRMISHIRI----- 1199
 QY 937 GSTNAVIVDAQGTGHHKRVRRSSFLNAKKLYEDQAKRVKOYLSNLEMEESLOT 996
 DB 1200 -----RQHQQTYY-----KIEQPKVTVQYLVDETVLDESLYE 1234
 QY 997 SLOCEPATNT 1006
 DB 1235 SLRIEPLKPT 1244

RESULT 11
 CC25_YEAST
 ID CC25_YEAST STANDARD; PRT; 1589 AA.
 AC P04821;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cell division control protein 25.
 GN CDC25 OR CTN1 OR YLR310C OR L2142.6.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87131091; PubMed=3545497;
 RA Broek D., Toda T., Michaeli T., Levin L., Birnmeier C., Zoller M.,
 RA Powers S., Wigler M.,
 RT "The S. cerevisiae CDC25 gene product regulates the Ras/adenylyate
 RT cyclase pathway."
 RL Cell 48:789-799(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86220116; PubMed=3011405;
 RA Camonis J.H., Kalekline M., Gondre B., Garreau H., Boy-Marcotte E.,
 RA Jaquet M.,
 RT "Characterization, cloning and sequence analysis of the CDC25 gene
 RT which controls the cyclic AMP level of Saccharomyces cerevisiae."
 RL EMBO J. 5:375-380(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,

```

RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Raich A., Trevaskis E., Vignati D.,
RA Wilcox L., Woldman P., Vaudin M., Wilson R., Waterston R.,
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP DOMAINS.
RX MEDLINE=89181526; PubMed=3070351;
RA Munder T., Mink M., Kuentzel H.;
RT "Domains of the Saccharomyces cerevisiae CDC25 gene controlling
RT mitosis and meiosis.";
RL Mol. Gen. Genet. 214:271-277(1988).
RN [5]
RP FUNCTION.
RX MEDLINE=91203884; PubMed=2017169;
RA Jones S., Vignats M.L., Broach J.R.;
RT "The CDC25 protein of Saccharomyces cerevisiae promotes exchange of
RT guanine nucleotides bound to ras.";
RL Mol. Cell. Biol. 11:2641-2646(1991).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,
CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03579: CAA27259.1; -
DR EMBL: M15458: AAA34478.1; -
DR EMBL: U17247: AAB67360.1; -
DR EMBL: U20618: AAB64528.1; -
DR PIR: A26596: RCHYCS.
DR HSSP: P00519: IABL.
DR SGD: S0004301: CDC25.
DR InterPro: IPR000651: RASGEFN.
DR InterPro: IPR001895: RASGEF_CDC25.
DR InterPro: IPR001452: SH3.
DR Pfam: PF00018: SH3; 1.
DR Pfam: PF00617: RASGEF; 1.
DR Pfam: PF00618: RASGEFN; 1.
DR PRINTS: PR00452: SH3DOMAIN.
DR PRODOM: PD000066: SH3; 1.
DR SMART: SM00147: RASGEF; 1.
DR SMART: SM00229: RASGEFN; 1.
DR SMART: SM00326: SH3; 1.
DR PROSITE: PS00720: GDS_CDC25; 1.
DR PROSITE: PS50002: SH3; 1.
KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;
KW Mitosis; Transmembrane; SH3 domain.
FT TRANSMEM 1452 1473 POTENTIAL.
FT DOMAIN 58 128 SH3.
FT CONFLICT 497 497 I -> Y (IN REF. 2).
FT CONFLICT 954 963 PUGHHEPKN -> LSTYMLSR (IN REF. 2).
SQ SEQUENCE 1589 AA; 179091 MW; CA90053435C85465 CRC64;

```

```

Query Match 3.7%; Score 282.5; DB 1; Length 1589;
Best Local Similarity 22.1%; Pred. No. 1,1e-05;
Matches 133; Conservative 108; Mismatches 244; Indels 117; Gaps 22;

```

```

OY 333 VLVANNHFNDECDPAWTRFL---EEFNNLEKMGKHLRLINAC----- 377
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 988 VLVASODEFFNNLE-----FLNNSDFKDACER-----YVEISNLCIIVDQLEEREN 1035
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 378 -----AAKARLRLMTKPSREAPLPIILLGSEKGFIF-VDSVSSGSKATEAGLKRQDQ 432
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1036 LNVYAAARMKNLTAE-----LTKGEQEKWFDIYSDYSDDESENDEAIT--DDE 1083
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

OY 433 ILEWNGONFENIOLSKAMEILRNTHLSTVKTNLFVFKELLRLSEKRNCAHPHPIKIG 492
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1084 L-----GSDDIYERKRAANEKN-----LPWLTSDYEISLVYDSGKRGSTKEA 1128
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 493 DIKASHYSIPDLAVDEQVIGLEKYNKSKANTVCGRNKKLILDKTRISIPQR--PY 550
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1129 LIEHLSISHELVDAAFNVTMLTFERSI-----LTTREFFALI--YRYNLTYPGGLST 1178
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 551 NDIGIGSODDSIVGLQRTGHIPTALPVSGTSSNPDLLQSHHRLDFAATDLPDQVL 610
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1179 DDNINWIEKSNPKCKVNMIMTFLQYTRNIYEPGI---PLLNTRAK----- 1225
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 611 RVFKADQOSRYIMISKDTAKEVYIAIRFEAVATAPDQYSLCEVSVPTEGVIKORLPD 670
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1226 -----WVSEKIPGAEDLLQKINCKLINENKE-----PYDP---KQD--- 1261
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 671 QLSKADRIOLSGRYIYKNNMEFETLCSDDAOELRESOISLQISTYEVATQLSMRNF 730
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1262 SVSAVQVTRKDNKSP--HMSSSSLPSSA--SSAFRLKRLKLLDDIDPYATATQVLLEH 1318
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 731 ELFRNTEPTETIDLP--KLRSKTSANLKRFEEDVINOEFVWVASFLRETNOLOKRMKII 788
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1319 DLYRITMEFCLDRAMQTKYCNMGSGPNITKFTANANTLTNFVSHITVQADVKTSKLT 1378
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 789 KHPKIALHCECKENFENFAIISGLNLAPARLRTWEKLPKRYEKLFDODLPDPSR 848
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1379 QYFVTVAHQHCKELNPFSSMAIYSAIYSSPIYRLKTKWDLVRESKDLKLNLMDSKR 1438
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 849 NMAKRYAVNLSONLQPIIPLFPYIKKDLFLHEGDS---KVDGLVFEKLMIAKEIR 905
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1439 NFKVYRELRLSVT--DVACVEFFGYLSDLTFTEVGNPDFLNSTNINFSKRTKIANIVE 1497
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 906 HV 907
DB 1498 EI 1499

```

```

RESULT 12
STEB6_SCHPO STANDARD; PRT; 911 AA.
ID STEB6_SCHPO
AC P26674; 09JUM8;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Steb6 protein.
GN STEB6 OR SPCCL450.17 OR SPCCL442.01.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_Taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90190870; PubMed=2107403;
RA Hughes D.A., Fukui Y., Yamamoto M.;
RT "Homologous activators of ras in fission and budding yeast.";
RL Nature 344:355-357(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Monev P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

```

```

RA Woodward J., Volckaert G., Aert R., Robben J., Grymompres B.,
RA Meljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Landt I., Beck A., Leinach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Mambuli R., Purcell B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaie V., Motter S.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpkovski G.V., Usero D., Barrell B.G., Nurse P.,
RA "The genome sequence of Schizosaccharomyces pombe.",
RL Nature 415:871-880(2002).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN IS ESSENTIAL FOR MATING.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53254; CAA37345.1; -
DR EMBL: AL049559; CAB40184.1; -
DR EMBL: AL031966; CAA21435.1; -
DR PIR: S28098; S28098.
DR InterPro: IPR000651; RasGEF.
DR InterPro: IPR001895; RasGEF_CDC25.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 1.
DR Pfam: PF00617; RasGEF; 1.
DR Pfam: PF00618; RasGEF; 1.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00147; RasGEF; 1.
DR SMART: SM00229; RasGEF; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PROSITE: PS50002; SH3; 1.
DR Guanine-nucleotide releasing factor; SH3 domain.
KW DOMAIN
FT SEQUENCE 911 AA; 105185 MW; 5F8C12D20C4B753F CRC64;
SQ
Query Match 3.6%; Score 277; DB 1; Length 911;
Best Local Similarity 28.0%; Pred. No. 8; 6e-06;
Matches 78; Conservative 49; Mismatches 122; Indels 30; Gaps 7;
OY 650 YSLCEVSVPEGVYKORLPDLSK-----ADRIQLSGRYLKNMNETLCSDEDA 702
DB 594 YKVEFEVVPVH-FASAEELHLSHLHPSTKSRHMEGK---ELSOLELDSLHNSP 649
OY 703 QELRESOISLQSLVEVATQISMRNFELRIETETIDYDFKLRSTSCNLRKEE 762
DB 650 DPLIYKDELVL-LPPREIAKQCLILEFQSFHSIRIQELTKTW-----NLNRFSD 700
OY 763 VINQETFW-----VASEILRETNLQKRMKILKHFIRIALCHRCDFNSMFATISGL 814
DB 701 KEKTSFYLSNHLVNFVETIYVGESEPRRTNVLAFIVQCDLRLLNPFASLFIISAL 760
OY 815 NLAPVARLRTWEKLEFNKYKEKLFQDLQDLFPDSRNMARRNVINSONLOPPIIFLPVVK 874
DB 761 NSSPIHRLKRTYANLNSKTLASFELNLLTEARKNFSNRYRDCI--ENCVLPCVPFLGYVF 818
OY 875 KDLTFLEHGNDKSVQGLVNPFEKLRMIKELRHGRMAV 913
DB 819 TDLTFLEKTKGNKFNQNMNINDKRTKVTRIINLEIKFQSV 857
RESULT 13

```

```

SOS2_MOUSE
ID SOS2_MOUSE STANDARD; PRT: 1297 AA.
AC 002384;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Son of sevenless protein homolog 2 (SOS-2) (mSOS-2) (Fragment).
GN SOS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Eye;
RX MEDLINE=9235328; PubMed=1631150;
RA Bowtell D.D., Fu P., Simon M.A., Senior P.V.;
RT "Identification of murine homologues of the Drosophila son of
RT sevenless gene: potential activators of ras.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST EMBRYONIC AND ADULT TISSUES.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z11664; CAA7732.1; -
DR PIR: S25714; S25714.
DR HSSP: O62245; 1PMS.
DR MGD: MGI:98355; SOS2.
DR InterPro: IPR001331; GDS_CDC24.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000651; RasGEF.
DR InterPro: IPR001895; RasGEF_CDC25.
DR InterPro: IPR000219; RhogEF.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00617; RasGEF; 1.
DR Pfam: PF00618; RasGEF; 1.
DR Pfam: PF00621; RhogEF; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00147; RasGEF; 1.
DR SMART: SM00229; RasGEF; 1.
DR SMART: SM00325; RhogEF; 1.
DR PROSITE: PS00741; DH_1; FALSE_NEG.
DR PROSITE: PS50010; DH_2; 1.
DR PROSITE: PS500720; GDS_CDC25; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR Guanine-nucleotide releasing factor.
KW NON_TER
FT DOMAIN 164 353 DH.
FT DOMAIN 407 510 PH.
FT DOMAIN 740 926 RAS-GEF.
FT DOMAIN 1145 1148 POLY-PRO.
FT DOMAIN 1164 1173 POLY-PRO.
SQ SEQUENCE 1297 AA; 148846 MW; 3A53F8515BF9D625 CRC64;
Query Match 3.2%; Score 245.5; DB 1; Length 1297;
Best Local Similarity 18.3%; Pred. No. 0.00042;
Matches 253; Conservative 188; Mismatches 467; Indels 477; Gaps 60;
OY 79 PETAVDSED-----DDDEEDIERAS--DPLMSRDIVRDLCKPDIRTDD 121
DB 29 PRTVQVEERVQKTFPHPIDKMAIDADAQSAIEKRKRNRNPLLT-----PVDKIH 77

```

```

QY 122 DIEOLLEPMHQLPAPAMTMSVRELCVAVFA---VVERAGTIVLNDGELDSWVILN 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 SLEKVLGY-----KVYHVSLLYAVLEYISADILKLAGNYVNRH----- 119
QY 179 GSVEVTVYPDGKAILCMGNSFGVSPMDKEYKGMKTKVYDCQFVCIADQYCR1--LN 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 --VEISQODIK-----VSMCADKVLMD--MFDQDDDIGVLSCDEDCSSGELN 164
QY 237 QVEKNMOKVEEGBE1-----VMKREHRE---LDRTGKRKHVYIKGSEKLTMLHVEHS 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 YDLVLTETLAEERQYLLNELMIIKVFREAPFLDLRKLKLPSEI-----EKIFSNISDIHE 218
QY 289 VVDPPT--FIDFELTYTFLFSSPMVEGCKLLEWNPDLSDKTYR-----VLLVMVNHFN 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 LTKKLGLIID--TVEVTDBSSPHPLAGSCFEDLAEQADRPETLSQDILLABEPNDFS 276
QY 343 DFECDPAMTRFLEFENNLERE-----KMGHLRLNT-----ACA 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 KLMARPAVALHFGSIADGFKFAVRVYLPRLMLVYVYCHMYFELLKCLKACSEQDEKCL 336
QY 379 AKAKRRLMTLTKSRERAPLRFILLGSEKGFIFVDSVDSGSKATEGLKRGDQ1--LEVN 437
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 337 NQATITLALNLO-----GSMRITYKQHSFRRRGDVCLEFYN 372
QY 438 GO--NFENIOLSKAMEILRN-----NTHLSIVKTNLFVFKELLT 475
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 ROLRSKRLAKKKMNEIOLKNDGMGKDIGOCNFEIMEGLTRIGAKHEHIFELFDLM 431
QY 476 RLSEKRNKAPHLR-----KIDIKKAS--RYSIFDLAVDEOYI-- 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 -ISCKPMHGOTRLRGYSSAEYRLKEKFMKRIQICDEKEDACEYHAEVLSKENSIFYA 490
QY 514 ---GLEVNNKSKANVTYGGNKKLILDKTRISILPKRPYNDIGGSDSIVGLQTK 570
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 AKSAEENNNMMAALISLHYSITLDRMD---SVL-----LKEEN 526
QY 571 HIPTALPVSGTSSNPDLQSHRILDFSATPDLPOVLRFVADQOSRY-1WISKDTT 629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 527 EQPLRLP-----SPDMWR-----FYVT-DSEENI--VEFDNKSISGPIITINGGT 568
QY 630 AKEVY-----IQAIRFAVTATPDDYSLCEVSTPREGVYKQRLRPOLSKLDRIOL-- 681
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 569 VVKLIERLTVMYADPNFVFTLTYSFCKPOELLMLIEREPERPEPTEADAKIALEK 628
QY 682 -----SGRY-LKNNME-----TELCS----- 698
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 629 GEOPISDLKRFKREYQVPOVLRYLVNFRHMHVHYDFEDELRLERLESFISVYGA 688
QY 699 -----DEDAQELLRESQ-----ISLLQSTIVEVA 722
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 689 MKKVESIAKIKRKKQOANGISHNITFESSPPVEMHISRTGOFTPLMTLHPIEIA 748
QY 723 TOLSMRNFELFRNIEPTVEYIDDLFKLSK--TSCANLKRFEVINOEFVVASSELRETNO 781
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 749 RQTLLESDDLRYKQVPSBELGVSWTKEDKEINSPLLMKIRHTTNLTLWMEKCIVEAENF 808
QY 782 LKRAKIKTHFKIALHCRECKNFNSMAFISGLNLAVALPRLTJRTMKYKXELFPDLO 841
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 809 EERAVLSRIYEILQVODLNNFNGVLEISAVSVSYRLDHFELQEKRRILDDAV 868
QY 842 DLDPSSRMMAKYRVNLNSQNLQPIPLFPVYIKKDLFLHEGNSD---KVDGLVNEKL 897
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 869 EL--SQOHFKKY--LVTKISINPCVPEFFGIYLTNIIKTEBGSNDFLKRGKDLINFSKR 924
QY 898 RMLAKETLRHVGMAVVMMDPALMFRTRKKKYR-----SLGSLSQ----- 936
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 925 RKVA-EI--TGEIOQYONOPYCL-RTEPEMRFFENINPMOILSEKFTDYLFNKSLTEI 980
QY 937 -----GSTNATVLDVAQGTGKKKRVARRSFLNAKK 966
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 981 PRNCKOPPRPRPKTFSLSKSGIRPNAGRGSTISGTL-----RGHPLPLEREPYKISFS 1034
QY 967 LYEDAMARKVKOYLS-----NLEMEDESLOTLSLOCEPATNTL----- 1007

```

```

Db 1035 RIAETLESTVASPTSPNTPSPVPGASSDHSVFLDVLDNLSGSGSNTIPAPVLLPHSKTF 1094
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1008 -----PKNPQDK-----PVKSETSPVAPRASQOKAOSLPP 1040
Db 1095 FSSCGSLHKLSEBPLIPPLPRPKKFDHALNSKGAVKSDDDPPA-----IP-P 1142
QY 1041 QOOPRAHKINOGLQY-----PAVSLYPSRKKVVPVKLPP-----FGINSPOALK 1085
Db 1143 KOPPPP--KVKPVVPLMGTFDGVPSPPPPPRPDLPPTPPVPLRPPEHFINCFNIO 1200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1086 KILSLESGSLERHKQOAEETJSMASQSLSPPTSPQSS--PRKGYTLADSGVDFNFSQSG 1144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1201 P-----PRLGHPHDPDLWRD--VSTCPNSPSTPTPTSPRIPRSCHLLS-----SS 1245
QY 1145 HSEIS 1149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1246 HSSILA 1250

RESULT 14
RGL1_HUMAN
ID RGL1_HUMAN STANDARD; PRT; 768 AA.
AC Q9NZL6; Q9NZL5; Q9HBY3; Q9HBY4; Q9Y2G6; Q9UG43;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ral guanine nucleotide dissociation stimulator-like 1 (RalGDS-like 1).
GN RGL1 OR RGL OR KIA00959.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
ON NCBI_Taxid=9606;
RX [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RX MEDLINE=20225471; PubMed=10760592;
RA Sood R., Makalowska I., Carpin J.D., Robbins C.M., Stephan D.A.,
RA Connors T.D., Morgendesser S.D., Su K., Pinkett H.W., Graham C.L.,
RA Quesenberry M.I., Baxevanis A.D., Klingner K.W., Trent J.M.,
RA Bonner T.I.;
RT "The human RGL (RalGDS-like) gene: cloning, expression analysis and
RT genomic organization.";
RL Biochim. Biophys. Acta 1491:285-288(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
RN [3]
RP SEQUENCE OF 15-768 FROM N.A. (ISOFORM A).
RC TISSUE=Brain;
RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE GUANINE NUCLEOTIDE EXCHANGE FACTOR.
CC -1- SUBUNIT: INTERACTS WITH RAS (BY SIMILARITY)
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES WITH
CC STRONG EXPRESSION BEING SEEN IN THE HEART, BRAIN, KIDNEY, SPLEEN
CC AND TESTIS.
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation in
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

or send an email to license@isb-sib.ch).

```
CC -----
DR EMBL: AF186779: AAF67280.1: -
DR EMBL: AF186780: AAF67281.1: -
DR EMBL: AF186798: AAG14400.1: -
DR EMBL: AF186781: AAG14400.1: JOINED.
DR EMBL: AF186783: AAG14400.1: JOINED.
DR EMBL: AF186784: AAG14400.1: JOINED.
DR EMBL: AF186785: AAG14400.1: JOINED.
DR EMBL: AF186786: AAG14400.1: JOINED.
DR EMBL: AF186787: AAG14400.1: JOINED.
DR EMBL: AF186788: AAG14400.1: JOINED.
DR EMBL: AF186789: AAG14400.1: JOINED.
DR EMBL: AF186790: AAG14400.1: JOINED.
DR EMBL: AF186791: AAG14400.1: JOINED.
DR EMBL: AF186792: AAG14400.1: JOINED.
DR EMBL: AF186793: AAG14400.1: JOINED.
DR EMBL: AF186794: AAG14400.1: JOINED.
DR EMBL: AF186795: AAG14400.1: JOINED.
DR EMBL: AF186796: AAG14400.1: JOINED.
DR EMBL: AF186797: AAG14400.1: JOINED.
DR EMBL: AF186798: AAG14401.1: JOINED.
DR EMBL: AF192520: AAG14401.1: JOINED.
DR EMBL: AF186783: AAG14401.1: JOINED.
DR EMBL: AF186785: AAG14401.1: JOINED.
DR EMBL: AF186786: AAG14401.1: JOINED.
DR EMBL: AF186787: AAG14401.1: JOINED.
DR EMBL: AF186788: AAG14401.1: JOINED.
DR EMBL: AF186789: AAG14401.1: JOINED.
DR EMBL: AF186790: AAG14401.1: JOINED.
DR EMBL: AF186791: AAG14401.1: JOINED.
DR EMBL: AF186792: AAG14401.1: JOINED.
DR EMBL: AF186793: AAG14401.1: JOINED.
DR EMBL: AF186794: AAG14401.1: JOINED.
DR EMBL: AF186795: AAG14401.1: JOINED.
DR EMBL: AF186796: AAG14401.1: JOINED.
DR EMBL: AF186797: AAG14401.1: JOINED.
DR EMBL: AB023176: BAA76803.1: ALF_INIT.
DR EMBL: AL080117: CAB45716.1: -
DR HSP: 003386: L1XD.
DR MIM: 60567: -
DR InterPro: IPR000159: RA_domain.
DR InterPro: IPR000651: RasGEF.
DR InterPro: IPR001895: RasGRF_CDC25.
DR Pfam: PF00617: RasGEF_1.
DR Pfam: PF00618: RasGEF_1.
DR Pfam: PF00784: RA_1.
DR SMART: SM00314: RA_1.
DR SMART: SM00147: RasGEF_1.
DR SMART: SM00229: RasGEF_1.
DR PROSITE: PS00720: GDS_CDC25_1.
KW Guanine-nucleotide releasing factor; Alternative splicing.
FT DOMAIN 236 499 RAS-GEF.
FT DOMAIN 541 613 SER-RICH.
FT DOMAIN 648 735 RBD.
FT VARSPIC 1 9 MLLIMQAKM -> MEYKPYGPTQEVSKTKLSTKVESTGHW
FT CONFLICT 15 15 LVEDHRLVEVLTKEE (IN ISOFORM B).
FT CONFLICT 687 687 W -> G (IN REF. 3).
FT CONFLICT 761 761 N -> Y (IN REF. 3).
SQ SEQUENCE 768 AA: 86700 MW: 5EEKAFGD5E0C2386 CRG64;
```

Query Match 3.2%: Score 244.5; DB 1: Length 768;
Best Local Similarity 20.7%: Pred. No. 0.00022;
Matches 170; Conservative 112; Mismatches 306; Indels 233; Gaps 31;

```
OY 605 LPOQVIRVKKADQSRVYIMSKDT-----TAKRVYQAIREFAVTATPPQYSICE----- 654
DB 77 LVENLTATF-GQNDFTYISITFISTYNGPASTKVEVLELLDVRGNLTSPN---CEEDGSG 131
OY 655 VSVTPGCVIKO-----RRLPDQ-----LSKLAD-----RIQLS 682
```

```
DB 132 SSSESKEVIRNAIASILRAMLDQCAEDFREPHFPCLQKLLDYLITRMMPGSDPERRAQN 191
OY 683 GRYYLKNMETER-----LCSDEDAQELLRESQIILLQSTVEVATQLSMRFFELPRNI 736
DB 192 LEQFOKEVEFTDGLPNTISFSLSEEELEGGSAEFTCFSEDLVABQLYMAQLEFKVY 251
OY 737 EPEVEYIDDLFKLSK-----TSCANLKRFEVINOETFWVASEL--RETNOQLRMK 786
DB 252 VPHHGCICMSRRDKKENKHLAFTIRATISQF-----NTLTKCVSVITLGGKELKTQORAK 307
OY 787 IIKHFILALHCRECKNENSMFAIISGLNLAPYARLRTTWKLPNKYEKLFQDLQDLDP 846
DB 308 IIEKWINIAHECRFLKNFSSLRATVSALOSNSYRLKTTAAVPRDMLMFEELSDIFSD 367
OY 847 SRMMAKYRVNLSQ-----NLQPL-----IPLPVYIKKL 877
DB 368 HNNHLTSRELLMKEGTSKPNLDSSVKNQKRTQRRLOLQDKMGVQGTVPYVIGTFLTDL 427
OY 878 TFLHEGNDGKVD-GLVNFELRMIAKEIRHYGRMASVNDMPALMFTRRKKMWSLSLSQ 936
DB 428 TMLDTALQDYIEGGLINFERR--REFEYIAQI-----KLLQ 462
OY 937 GSTNATVLDVAQTGCHKRRYRSSFNAKKLYEDAQMARKVKOYLNLLEMDSESLQTL 996
DB 463 SACSNSYCMTPDQ-----KFIFMPQROLLEES-YAL 494
OY 997 SLOCEPATNTLPKNPGDKKPKVSETSPVAPRAGSQOKAOSLPQOQPPPAHKINQLOY 1056
DB 495 SCEIEAADAATSTSPKPRKSMVRLSLF--LQSDMTISFTPT--TKRDPKSTAGSGSESM 551
OY 1057 PAVSL--YPSRKVPVVKDLPFGINSPOALKILSLSEEGSLERHKKQAEDEITSNASSQ 1113
DB 552 DSVSVSCENSHSEAEGETTP--MDTPPEPKKLEBSSSCSSIH---SMDTNSGMS 606
OY 1114 LSSPPTSPQSSPRKGYTLAPSGTVDNFSDGSEISRSSTVSNSPSPVSLHDERQ 1173
DB 607 LNPPLSSPPSC-----NNPKIKHRSVSVSISITVLPYPYNOQNE 648
OY 1174 RHVSVI-VETNLGMGMERTTLEPDQYSLGSIAPMSEGGLYATATVYSSPTEELSD 1232
DB 649 TCIIIRISVEDNNG-----NMKRSIMLTNSQDKTPAVIOR 681
OY 1233 OCGRASIDLAADSGRGSWTSSGSHDNIOTIQHRSMETLPCFHTHFDYSGDPAGLW--- 1289
DB 682 AMLKHLND-----SDPAEEYELVOYISDEKEL-VIP-----DSANFYAM 720
OY 1290 ASSSHMDQIMFSDHSTKYNRONOSRES--LEQAOSRASMAS 1328
DB 721 NSQVNFDFILRKKNMSMEQYKLSRSLTLPRTAKKGCSN 761
```

RESULT 15
SOS2_HUMAN STANDARD; PRT; 1332 AA.
ID AC 007890: Q15503:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Son of sevenless protein homolog 2 (SOS-2).
GN SOS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93262494; PubMed=8493579;
RA Chardin P., Camonis J.H., Gale N.W., van Aelst L., Wigler M.H.,
RA Bar-Sagi D.;
RT "Human Sos1: a guanine nucleotide exchange factor for Ras that binds
to GRB2.";
RL Science 260:1338-1343(1993).

[2]
RN SEQUENCE OF 724-1296 FROM N.A.
RP TISSUE=Placenta;
RC Fath I., Apioi F., Dutrillaux B., Tocque B.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L3858; AAA35914.1; -;
CC EMBL: L20686; AAA91852.1; -;
CC HSSP: Q62245; 1PMS.
CC Genew: HGNC:11188; SOS2.
CC MIM: 601247; -;
CC InterPro: IPR001331; GDS_CDC24.
CC InterPro: IPR001849; PH.
CC InterPro: IPR000651; RASGEFN.
CC InterPro: IPR001895; RASGEFN.
CC InterPro: IPR000219; RHOGEF.
CC Pfam: PF00169; PH; 1.
CC Pfam: PF00617; RASGEFN; 1.
CC Pfam: PF00618; RASGEFN; 1.
CC Pfam: PF00621; RHOGEF; 1.
CC SMART: SM00233; PH; 1.
CC SMART: SM00233; PH; 1.
CC SMART: SM00147; RASGEF; 1.
CC SMART: SM00229; RASGEFN; 1.
CC SMART: SM00325; RHOGEF; 1.
CC PROSITE: PS00741; DH.2; 1.
CC PROSITE: PS00720; GDS_CDC25; 1.
CC PROSITE: PS50010; DH.2; 1.
CC PROSITE: PS50003; PH_DOMAIN; 1.
CC KW Guanine-nucleotide releasing factor.
CC -----
CC FT DOMAIN 198 388
CC FT DOMAIN 442 546
CC FT DOMAIN 775 961
CC FT DOMAIN 755 758
CC FT DOMAIN 1180 1183
CC FT DOMAIN 1203 1208
CC FT DOMAIN 778 778
CC FT CONFLICT 861 861
CC FT CONFLICT 948 948
CC FT CONFLICT 999 999
CC FT CONFLICT 1032 1032
CC FT CONFLICT 1042 1042
CC FT CONFLICT 1112 1114
CC FT CONFLICT 1159 1159
CC FT CONFLICT 1296 1296
CC SEQUENCE 1332 AA; 153030 MW; EF4BA9B52FA85AC9 CRC64;
Query Match 3.18; Score 242.5; DB 1; Length 1332;
Best Local Similarity 17.08; Pred. No. 0.00061;
Matches 220; Conservative 135; Mismatches 322; Indels 615; Gaps 50;

Db 458 HIFLFDGLMISCKRNHGRLPGYTSAEYRLKEKFMKIQICDKEDNCEHKAHELVS 517
QY 229 QDYCRLL---NOVEKN-----MOKVEEGEIVAVKREHL----- 260
Db 518 DENSIIFAAKSAEKKNNMAALISLHYRSTLDMDSVLKKEENQPLRLPSEYRYEV 577
QY 261 -----DRTGRKHIVIKTS-----EELTMHLEESVVDPTFIEFLTYRTE 305
Db 578 KQSEENVIVEDNLOSRSGLPIIKGTIVKILERLYHM-----YADPNVRVFLTYRSE 632
QY 306 LSSPMVEVKKLEWFN---DPSLRDK-----VTRVY 333
Db 633 -CKQPELSTLIRFRIPEPEPTDADKLAIEKEGEPIADLKRREKYQVQYLRANF 691
QY 334 LLMVNNHFNDFEDPAMTRELFEFENNLREKKGHLRLNTACAKAKRLMTLTKPSR 393
Db 692 RHVVDHYHDFEDLLELLELSEFISVR-----GKAKK----- 726
QY 394 EADLPFLLGSEKGFGEFVDSVDSGSKATEAGLRGQDILEVNGQNFENIOLSKAMEIL 453
Db 727 -----WVESI----- 731
QY 454 RNNTHLSTVKTNLVEYKELLIRLSEKRNAGAPHLKIDIKKASRYSLPDLAVDEQYI 513
Db 732 -----AKIIRKKQAOANGVSH----- 748
QY 514 GLEKVNKKSKANTVGGRNKLKILDKTRISILPQKRYNDIGISQSDDSIVGLRQTKHIP 573
Db 749 -----NITFESP----- 755
QY 574 TALPVSGTLSSNPDLLOSIIHRLIDFSAPPDLPDQYLRYFKAKDOOSRYIMSKDTAKEV 633
Db 756 -----PPIEMH-----ISK----- 765
QY 634 VIOAIRFEAVTAPPDQYSLCEVSTPEGVTKORRLPDQSLKADRIQLSGRYLKNNMET 693
Db 766 -----PGQF----- 769
QY 694 ETLCSDEDAQELRESQISLLOLSTVEVATOLSMRNEFLERNIEPTYEIDLFKLRSK-T 752
Db 770 ET-----FDLMTDPIELIAQQLLLESDDLRYKQPSBELVGSVTKRDKEL 814
QY 753 SCANLKRFEENVINOETFWVASEILRETNLQKRKKIKHFKIALHCRCKENSMFAIIS 812
Db 815 NSPMLKMIHTTNLTLMFEKCIYEAENFEERVAVLSRIEILQVQDNLNENGVALEIYS 874
QY 813 GNLAPYARLRTIWEKLPKNYEKLFODLOLDPSPNNMAKYRVNLVSQNLQPIIPLFPV 872
Db 875 AVNSVSYYRLDHTFEALQERRKRILOEAVEL--SDHFKKY--LVKLKSLNPPCVPEFGI 930
QY 873 IKKDLTFHEGNS---KVDGLVNFCKLRMAK---EIRHVG-----BMA5V 913
Db 931 YLNIILTEEGNNDPLKRRGKDLINFSKRRKVAEITGEIQYOYONQYCLRIEPMARFE 990
QY 914 NMDP-----ALMEFTRKKRW-----RSLGSUS-----QGSTNA 941
Db 991 NLNMGASAEKEFDYLLFNKSLIEIPRNCQPPRPFRKSTFLSKGIRPNTGHSSTG 1050
QY 942 TVLDVAOTGGHKKRVRS---SFLNAKKLYEDAQAMARKYQVLSNLELMEDESLQTS 997
Db 1051 TL-----RGHPPLRLEREPKISF-----SRIATELE-----STVS 1081
QY 998 LOCEPATNTLPKPNQDKPKPKSETS-----SRIATELE-----STVS 1081
Db 1082 APTSPNTPTSP-----PVSASDLSVFLVDVNLSSGNSLFAFVLLPHSKSFSSCG 1134
QY 1023 -----PVADR-----AGSQKASLPO-PQOQPPRAKINQGLQVPA 1058
Db 1135 SLNKLSEPLPLPLPRKRFHDHASNKGNMSDDDDPAIPROPPPP-KVAPRVAPVT 1193
QY 1059 -----VSLYPSRKKYVPVADLP-----FGINSQALKIILSSESGSLERHKKQA 1103
Db 1194 GAFDGPLHSPPPPRDLDPDTPPVPLRPPEHFINCPFNLDP---PVLGHLRSDWL 1249

Wed Feb 19 16:11:35 2003

us-09-911-826a-2.rsp

Page 17

```

Qy 1104 EDTISNASSQLSSPPTSPSS -PRGTYLAPS 1134
      | | : | : | : | | : |
Db 1250 RD-1STCPSNPSTPSTPSRPRVRCVYLSS 1280

```

Search completed: February 18, 2003, 21:24:14
Job time : 61 secs

BLANK PAGE